



TRABALHO DE CONCLUSÃO DE CURSO

**ANÁLISE DAS PERDAS TÉCNICAS NAS CONCESSIONÁRIAS DE
DISTRIBUIÇÃO DE ENERGIA ELÉTRICA DO BRASIL –
CLUSTERIZAÇÃO E IDENTIFICAÇÃO DAS VARIÁVEIS MAIS
SIGNIFICATIVAS.**

Louise Cugula Couto

Orientador: Jorge Andrés Cormane Angarita

Co-orientador: Anésio de Leles Ferreira Filho

Brasília, Novembro de 2017

UNIVERSIDADE DE BRASÍLIA

FACULDADE DE TECNOLOGIA

DEPARTAMENTO DE ENGENHARIA ELÉTRICA

Louise Cugula Couto

ANÁLISE DAS PERDAS TÉCNICAS NAS CONCESSIONÁRIAS DE
DISTRIBUIÇÃO DE ENERGIA ELÉTRICA DO BRASIL –
CLUSTERIZAÇÃO E IDENTIFICAÇÃO DAS VARIÁVEIS MAIS
SIGNIFICATIVAS.

Trabalho de Conclusão de Curso submetida ao
curso de graduação em Engenharia Elétrica da
Universidade de Brasília, como requisito
parcial para obtenção do Título de Engenheira
Eletricista.

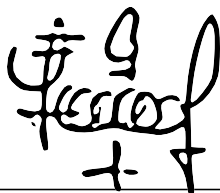
Orientador: Jorge Andrés Cormane Angarita
Co-orientador: Anésio de Leles Ferreira Filho

Brasília, Novembro de 2017

**ANÁLISE DAS PERDAS TÉCNICAS NAS CONCESSIONÁRIAS DE
DISTRIBUIÇÃO DE ENERGIA ELÉTRICA DO BRASIL –
CLUSTERIZAÇÃO E IDENTIFICAÇÃO DAS VARIÁVEIS MAIS
SIGNIFICATIVAS.**

LOUISE CUGULA COUTO

Trabalho de Conclusão de Curso submetido ao Departamento de Engenharia Elétrica da Faculdade de Tecnologia da Universidade de Brasília, como requisito parcial para obtenção do Título de Engenheira Eletricista, apresentada e aprovada em 28/11/2017 pela banca examinadora abaixo assinada:



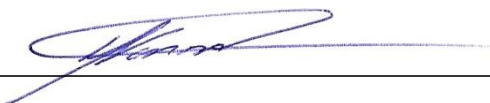
Prof.: Jorge Andrés Cormane Angarita, UnB

Orientador



Prof. Doutor: Anésio de Leles Ferreira Filho, UnB

Co-orientador



Mestre: Marcos Venícius Leite Vasconcelos, ANEEL

Membro Convidado

Brasília, Novembro de 2017

FICHA CATALOGRÁFICA

COUTO, LOUISE CUGULA

Análise das Perdas Técnicas nas Concessionárias de Distribuição de Energia Elétrica do Brasil - Clusterização e Identificação das Variáveis Mais Significativas [Distrito Federal] 2017.

81p., 210 x 297 mm (ENE/FT/UnB, Graduação, Trabalho de Conclusão de Curso - Universidade de Brasília, Faculdade de Tecnologia).

Departamento de Engenharia Elétrica

- | | |
|-------------------------------|-------------------------------------|
| 1. Perdas Técnicas de Energia | 2. Distribuição de Energia Elétrica |
| 3. Clusterização | 4. Regressão Multivariada |
| I. ENE/FT/UnB | II. Título (série) |

REFERÊNCIA BIBLIOGRÁFICA

COUTO, LOUISE C. (2017). Análise das Perdas Técnicas nas Concessionárias de Distribuição de Energia Elétrica do Brasil - Clusterização e Identificação das Variáveis Mais Significativas. Trabalho de Conclusão de Curso em Engenharia Elétrica, 2017, Departamento de Engenharia Elétrica, Universidade de Brasília, DF, 81p.

SESSÃO DE DIREITOS

AUTORA: Louise Cugula Couto

TÍTULO: Análise das Perdas Técnicas nas Concessionárias de Distribuição de Energia Elétrica do Brasil - Clusterização e Identificação das Variáveis Mais Significativas

GRAU: Engenheira Eletricista

ANO: 2017

É concedida à Universidade de Brasília permissão para reproduzir cópias deste trabalho de conclusão de curso e para emprestar ou vender tais cópias somente para propósitos acadêmicos e científicos. A autora reserva outros direitos de publicação e nenhuma parte desse trabalho de conclusão pode ser reproduzida sem autorização por escrito da autora.

*Dedico este trabalho à minha mãe, por quem nutro os mais profundos e verdadeiros
sentimentos de Amor e Gratidão.*

AGRADECIMENTOS

Agradeço primeiramente à minha família. O suporte e carinho incondicionais que minha mãe, Liliane, me proporcionou ao longo de toda minha trajetória acadêmica e pessoal foram peças fundamentais para que eu chegasse até aqui. Agradeço também à minha irmã, Christy Anne, cujos sábios conselhos sempre me recolocavam nos trilhos para que eu continuasse seguindo meus sonhos mesmo em momentos de desânimo ou preocupação.

Agradeço ao meu namorado, André, por sua compreensão e auxílio nas diversas fases da minha graduação. Seu amor, apoio e ombro amigo tornaram esta caminhada indubitavelmente mais suave.

Agradeço aos meus colegas de curso, Mariana, Eduardo e Pedro, que trilharam os passos da Engenharia comigo desde nosso primeiro dia na Universidade e com quem muito aprendi nos últimos anos. Sou muito grata por todos os momentos, desafios e conversas que compartilhamos.

Agradeço aos colegas da SRD/ANEEL, que tanto me ensinaram durante meu estágio na instituição. O suporte dado especialmente pelos colegas Marcos, Djane, Wesley, Davi Vidal e Leonardo foi essencial para minha evolução profissional e para a conclusão deste trabalho.

Por fim, agradeço aos meus mestres: professores e professoras que me ensinaram desde o alfabeto, soma e subtração às teorias mais complexas da Engenharia Elétrica. O conhecimento que hoje possuo é fruto de muito trabalho e dedicação dessas pessoas, que admiro e respeito profundamente. Em especial, agradeço ao professor Anésio, cujos acompanhamento e ensino nos últimos dois anos de minha graduação me trouxeram até aqui.

RESUMO

Perdas de energia são características intrínsecas aos sistemas elétricos de potência e ocorrem em todas suas etapas: geração, transmissão e distribuição. Dados do *World Energy Council* mostram que o nível de perdas de energia elétrica do Brasil encontram-se superiores às médias global e latinoamericana. A Agência Nacional de Energia Elétrica aponta, ainda, que somente o nível médio de perda de distribuição no Brasil gera um custo anual superior a R\$ 10,5 bilhões no setor. A determinação do nível dessas perdas mostra-se um desafio para os agentes do setor elétrico, especialmente devido às peculiaridades das redes de distribuição geridas por diferentes empresas. Nesse sentido, este trabalho busca identificar semelhanças e diferenças entre as concessionárias de distribuição brasileiras no que tange a gestão de perdas técnicas. Para tanto, são aplicados métodos de clusterização para o agrupamento das concessionárias em atuação no Brasil de acordo com seus níveis de perdas técnicas nos Sistemas de Média e Baixa Tensão (SDMT e SDBT). A análise de *clusters* visa segregar as empresas de distribuição com base em sua eficiência no que tange perdas técnicas de energia para posterior avaliação via regressão multivariada dos principais causadores dessas perdas no país. Os resultados indicam que, para as distribuidoras com menor nível de perdas na distribuição, os fatores mais influentes nos seus percentuais de perdas técnicas são a densidade de potência por quilômetro de rede em baixa tensão e a resistência média do segmento no mesmo nível de tensão. Por outro lado, as concessionárias com níveis medianos de perdas possuem como maiores motivadores para os resultados de perdas as densidades de energia fornecida por unidade consumidora por ano e de reguladores por quilômetro de rede em média tensão. Este último atributo também mostra-se expressivo nos resultados de perdas técnicas em SDMT e SDBT das empresas de distribuição com os piores índices de eficiência relativa a perdas técnicas, ao lado da densidade de potência nominal instalada em transformadores de distribuição por área atendida.

Palavras-chave: Eficiência na Distribuição, Perdas Técnicas, Clusterização, Regressão Multivariada

ABSTRACT

Energy losses are intrinsic to electrical power systems and occur in all its stages: generation, transmission and distribution. Data from the World Energy Council show that the level of electricity losses in Brazil is higher than the global and Latin American averages. The National Electric Energy Agency also points out that only the average level of distribution losses in Brazil generates an annual cost of more than R\$ 10,5 billion in the sector. The determination of the efficient level of these losses is a challenge for the electric sector agents, especially due to the peculiarities of the distribution networks managed by different companies. In this sense, this work aims to identify similarities and differences between Brazilian distributors in terms of technical losses management. Therefore, the study of clustering methods is proposed for grouping the companies that operate in Brazil according to their levels of technical losses in Medium and Low Voltage Systems. This cluster analysis aims to segregate the distribution companies based on their efficiency regarding technical losses for later evaluation through multivariate regression of the main motivators of these losses in the country. The results show that, for the distributors with the lowest level of technical losses in distribution, the most influential factors in their percentages of technical losses are the low voltage power density per network kilometer and the average resistance of the segment at the same voltage level. On the other hand, the companies with medium loss levels have as main motivators for their losses the density of energy supplied per consumer per year and the density of regulators per kilometer of medium voltage network. This last attribute is also significant in the results of technical losses for the distribution companies with the worst indices of efficiency regarding technical losses in the sector, along with the nominal power density installed in distribution transformers by served area.

Keywords: Distribution Efficiency, Technical Losses, Clustering, Multivariate Regression

LISTA DE FIGURAS

Figura 1.1 - Consumo final de energia por fonte.....	18
Figura 1.2 - Mapa do SIN 2017.....	20
Figura 1.3 - Mapa das concessões de distribuição de energia elétrica no Brasil.....	21
Figura 2.1 - Exemplo de dendrograma para clusterização hierárquica.	33
Figura 4.1 - Dendrograma da clusterização com base no percentual de perdas técnicas em níveis de média e baixa tensão.	54
Figura 4.2 - Evolução do parâmetro SR^2 em relação à quantidade de <i>clusters</i> formados no processo de agrupamento.....	55

LISTA DE TABELAS

Tabela 1.1 - Unidades Geradoras no Brasil e Capacidade de Geração.	19
Tabela 1.2 - Percentuais de perdas na transmissão e distribuição para diversos países entre 2000 e 2014.	22
Tabela 2.1 - Tipologias de rede em SDBT definidas pela ANEEL.....	28
Tabela 3.1 - Dados utilizados para a avaliação das concessionárias de distribuição brasileiras	43
Tabela 3.2 - Concessionárias analisadas.....	44
Tabela 4.1 - Percentuais de perdas técnicas em MT e BT das 60 distribuidoras analisadas	49
Tabela 4.2 - R^2 e SR^2 da clusterização das 60 distribuidoras com base no percentual de perdas técnicas em níveis MT e BT – Algoritmo de média das distâncias.....	51
Tabela 4.3 - R^2 e SR^2 da clusterização das 60 distribuidoras com base no percentual de perdas técnicas em níveis MT e BT – Algoritmo de distância dos centroides.	52
Tabela 4.4 - R^2 e SR^2 da clusterização das 60 distribuidoras com base no percentual de perdas técnicas em níveis MT e BT – Algoritmo de <i>Ward</i>	52
Tabela 4.5 - Grupos formados na clusterização das concessionárias com base no percentual de perdas técnicas em MT e BT.	56
Tabela 4.6 - Variáveis resmanescentes na análise de seleção com as respectivas transformações.....	57
Tabela 4.7 - Correlação de Pearson entre as variáveis independentes e a dependente para o <i>cluster</i> 1.	58
Tabela 4.8 - Correlação de Pearson entre as variáveis independentes e a dependente para o <i>cluster</i> 2.	59
Tabela 4.9 - Correlação de Pearson entre as variáveis independentes e a dependente para o <i>cluster</i> 3.	59
Tabela 4.10 - Variáveis independentes com correlações elevadas entre si e representativas de uma grandeza semelhante no <i>cluster</i> 1.	60

Tabela 4.11 - Variáveis independentes com correlações elevadas entre si e representativas de uma grandeza semelhante no <i>cluster</i> 2.	60
Tabela 4.12 - Variáveis independentes com correlações elevadas entre si e representativas de uma grandeza semelhante no <i>cluster</i> 3.	61
Tabela 4.13 - Variáveis dependentes remanescentes no modelo do <i>cluster</i> 1 após análise de correlação.	61
Tabela 4.14 - Variáveis dependentes remanescentes no modelo do <i>cluster</i> 2 após análise de correlação.	62
Tabela 4.15 - Variáveis dependentes remanescentes no modelo do <i>cluster</i> 3 após análise de correlação.	62
Tabela 4.16 - Variáveis retiradas do modelo por apresentarem VIF maior que 5 no <i>cluster</i> 1.	63
Tabela 4.17 - Variáveis retiradas do modelo por apresentarem VIF maior que 5 no <i>cluster</i> 2.	63
Tabela 4.18 - Variáveis retiradas do modelo por apresentarem VIF maior que 5 no <i>cluster</i> 3.	63
Tabela 4.19 - Variáveis utilizadas na regressão <i>stepwise</i> para o <i>cluster</i> 1.	64
Tabela 4.20 - Variáveis utilizadas na regressão <i>stepwise</i> para o <i>cluster</i> 2.	64
Tabela 4.21 - Variáveis utilizadas na regressão <i>stepwise</i> para o <i>cluster</i> 3.	64
Tabela 4.22 - Etapas da seleção <i>stepwise</i> para o <i>cluster</i> 1.	65
Tabela 4.23 - Etapas da seleção <i>stepwise</i> para o <i>cluster</i> 2.	65
Tabela 4.24 - Etapas da seleção <i>stepwise</i> para o <i>cluster</i> 3.	65
Tabela 4.25 - Médias aritméticas do percentual de perdas técnicas em níveis MT e BT e de cada um dos parâmetros mais explicativos para os <i>clusters</i> 1, 2 e 3.	66

LISTA DE ABREVIATURAS E SIGLAS

ABNT	Associação Brasileira de Normas Técnicas
AES-SUL	Aes Sul Distribuidora Gaúcha de Energia S/A
AME	Amazonas Distribuidora de Energia S/A
AMPLA	Ampla Energia e Serviços S/A
ANEEL	Agência Nacional de Energia Elétrica
AREAT	Área Total
BANDEIRANTE	Bandeirante Energia S/A
BOA VISTA	Boa Vista Energia S/A
BT	Baixa Tensão
CAIUÁ	Caiuá Distribuição de Energia S/A
CCEE	Câmara de Comercialização de Energia Elétrica
CEAL	Companhia Energética de Alagoas
CEB	Ceb Distribuição S/A
CEEE	Companhia Estadual de Distribuição de Energia Elétrica
CELESC	Celesc Distribuição S.A.
CELG	Celg Distribuição S.A.
CELPA	Centrais Elétricas do Pará S/A
CELPE	Companhia Energética de Pernambuco
CEMAR	Companhia Energética do Maranhão
CEMIG	Cemig Distribuição S/A
CEPISA	Companhia Energética do Piauí
CERON	Centrais Elétricas de Rondônia S/A
CFLO	Companhia Força e Luz do Oeste
CGH	Central Geradora Hidrelétrica
CHESP	Companhia Hidroelétrica São Patrício
CNEE	Companhia Nacional de Energia Elétrica
COCEL	Companhia Campolarguense de Energia
COELBA	Companhia de Eletricidade do Estado da Bahia
COELCE	Companhia Energética do Ceará
CompTotMed_B	Comprimento Total Médio

CompTotMed_MT	Comprimento Total Médio MT
COOPERALIANÇA	Cooperativa Aliança
COPEL	Copel Distribuição S/A
COSERN	Companhia Energética do Rio Grande do Norte
CP	Coeficiente de Perdas
CPFL LESTEPAULISTA	Companhia Leste Paulista de Energia
CPFL PIRATININGA	Companhia Piratininga de Força e Luz
CPFL SULPAULISTA	Companhia Sul Paulista de Energia
CPFL JAGUARI	Companhia Jaguari de Energia
CPFL MOCOCA	Companhia Luz e Força de Mococa
CPFL PAULISTA	Companhia Paulista de Força e Luz
CPFL SANTACRUZ	Companhia Luz e Força Santa Cruz
DEC	Duração Equivalente de Interrupção por Unidade Consumidora
DEMEI	Departamento Municipal de Energia de Ijuí
Dens_Regul_Area	Densidade de Reguladores por Área
Dens_Regul_Rede	Densidade de Reguladores por km de rede MT
Dens_Regul_UC	Densidade de Reguladores por UC
Dens_SInstTD_Area	Densidade de Potência Nominal Instalada de TD por Área
Dens_SInstTD_Rede	Densidade de Potência Instalada de TD por km de Rede MT e BT
Dens_SInstTD_UC	Densidade de Potência Instalada de TD por UC
Dens_TD_Area	Densidade de TD por Área
Dens_TD_Rede	Densidade de TD por km de Rede MT e BT
Dens_TD_UC	Densidade de TD por UC
Dens_UC_Area	Densidade de Unidades Consumidoras por Área
Dens_UC_Rede	Densidade de Unidades Consumidoras por km de Rede MT e BT
DensEnerg_Forn_Area	Densidade de Energia Fornecida Anual por Área
DensEnerg_Forn_Rede	Densidade de Energia Fornecida Anual por km de Rede MT e BT
DensEnerg_Forn_UC	Densidade de Energia Fornecida Anual por UC
DensPot_B_Rede	Densidade de Potência por km de rede BT
DensPot_MT_Rede	Densidade de Potência por km de rede MT

DensRed_B_Area	Densidade de Rede BT por Área
DensRede_MT_Area	Densidade de Rede MT por Área
DMED	Dme Distribuição S.A
EBO	Energisa Borborema Distribuidora de Energia S.A.
EDEVP	Empresa de Distribuição de Energia Vale Paranapanem
EEB	Empresa Elétrica Bragantina S/A
EFLJC	Empresa Força e Luz João Cesa Ltda
EFLUL	Empresa Força e Luz Urussanga Ltda
ELEKTRO	Elektro Eletricidade e Serviços S/A
ELETROACRE	Companhia de Eletricidade do Acre
ELETROCAR	Centrais Elétricas de Carazinho S/A
ELETROPAULO	Eletropaulo Metropolitana Eletricidade de São Paulo S/A
ELFSM	Empresa Luz e Força Santa Maria S/A
EMG	Energisa Minas Gerais - Distribuidora de Energia S.A.
EMS	Energisa Mato Grosso do Sul - Distribuidora de Energia S.A.
EMT	Energisa Mato Grosso - Distribuidora de Energia S.A
EnerTot_Forn	Energia Total Fornecida
EnerTot_Inj	Energia Total Injetada
EnerTot_MTBT	Energia Total Injetada em MT e BT
ENF	Energisa Nova Friburgo - Distribuidora de Energia S.A
EOL	Usina Eólica
EPB	Energisa Paraíba - Distribuidora de Energia S.A.
EPE	Empresa de Pesquisa Energética
ESCELSA	Espírito Santo Centrais Elétricas S/A
ESE	Energisa Sergipe - Distribuidora de Energia S.A.
ETO	Energisa Tocantins - Distribuidora de Energia S.A.
FEC	Frequência Equivalente de Interrupção por Unidade Consumidora
HIDROPAN	Hidroelétrica Panambi S/A
IENERGIA	Iguaçu Distribuidora de Energia Elétrica Ltda
LIGHT	Light Serviços de Eletricidade S/A
MT	Média Tensão
MUX	Muxfeldt Marin & Cia. Ltda

N_CircB	Número de circuitos BT
N_CircMT	Número de circuitos MT
N_Regul	Número de Reguladores
N_TD	Número de TD
NUC	Número de Unidades Consumidoras
ONS	Operador Nacional do Sistema Elétrico
P_MedTot_B	Potência Média Total BT
P_MedTot_MT	Potência Média Total MT
PCH	Pequena Central Hidrelétrica
Perc_SInst_TD	Percentual de Potência Nominal Instalada em TD
PercPerdas_MTB	Percentual de Perdas MT e BT
PercRede_B	Percentual de Rede B na Distribuidora
PercRede_MT	Percentual de Rede MT na Distribuidora
PRODIST	Procedimentos de Distribuição
PRORET	Procedimentos de Regulação Tarifária
ResMed_Ramal_B	Resistência Média do Condutor Ramal BT
ResMed_Segm_B	Resistência Média dos Segmentos de Rede BT
ResMed_Segm_MT	Resistência Média dos Segmentos de Rede MT
RGE	Rio Grande Energia S/A
S_InstMed_Reg	Potência Instalada Média por Regulador
S_InstMed_TD	Potência Instalada Média por TD
S_InstTot_Regul	Potência Instalada Total de Reguladores
S_InstTot_TD	Potência Aparente Instalada de TD Total
SDAT	Sistema de Distribuição em Alta Tensão
SDBT	Sistema de Distribuição em Baixa Tensão
SDMT	Sistema de Distribuição em Média Tensão
SIN	Sistema Interligado Nacional
SULGIPE	Companhia Sul Sergipana de Eletricidade
TD	Transformador de Distribuição
UC	Unidade Consumidora
UFV	Usina Fotovoltaica
UHE	Usina Hidrelétrica
UHENPAL	Usina Hidroelétrica Nova Palma Ltda

UTE	Usina Termelétrica
UtilizMed_Regul	Percentual Médio de Utilização de Reguladores
UtilizMed_TD	Percentual Médio de Utilização de TD
UTN	Usina Termonuclear
VIF	<i>Variance Inflation Factor</i>

LISTA DE SÍMBOLOS

ΔE_{RM}	Perdas de Energia para os Ramais de Ligação e Medidores
ΔE_{SDBT}	Perdas de Energia em SDBT
ΔE_{SDMT}	Perdas de Energia em SDMT
ΔE_T	Perdas de Energia para os Transformadores
Δp_{cu}^t	Perda de Potência para a Demanda Média no Cobre do Transformador t
Δp_{fe}^t	Perda de Potência no Ferro ou a Vazio do Transformador t
ΔP_M	Perda de Potência na Bobina de Tensão do Medidor
ΔP_R	Perda de Potência para Demanda Média em Ramais de Ligação
ΔP_{SDBT}	Perda de Potência para a Demanda Média em SDBT
ΔP_{SDMT}	Perda de Potência para a Demanda Média em SDMT
ΔP_{TR}	Perda de Potência para a Demanda Média em Transformadores
ΔT	Período de Tempo em Análise
A	Ampère
$\cos\varphi$	fator de potência
CP_R	Coeficiente de Perdas do Ramal
CP_{SDBT}	Coeficiente de Perdas no SDBT
CP_{SDMT}	Coeficiente de Perdas em SDMT
CP_T	Coeficiente de Perdas do Transformador
CT^{SDMT}	Comprimento Total do Trecho em SDMT
CT^{TRECHO}	Comprimento do Trecho classificado como Tronco
E^{SDMT}	Energia Fornecida pelo Alimentador em SDMT
F_d	Fator de Diversidade
i	Densidade de Corrente
I_j	Corrente Total a Jusante do Trecho Elementar
I_m^f	Corrente Média de Fase

I^{SDMT}	Corrente Média no Alimentador em SDMT
kV	Quilovolt
kW	Quilowatt
l	Comprimento do Trecho Elementar
L_{circ}	Comprimento Total do Circuito em SDBT
MW	Megawatt
MWh	Megawatt-hora
N_{UC}^1	Número de Unidades Consumidoras Alimentadas em 1 Fases e 2 Fios
N_{UC}^2	Número de Unidades Consumidoras Alimentadas em 2 Fases e 3 Fios
$N_{\text{UC}}^{2'}$	Número de Unidades Consumidoras Alimentadas em 1 Fases e 3 Fios
N_{UC}^3	Número de Unidades Consumidoras Alimentadas em 3 Fases e 4 Fios
P_{C}	Perda por Circuito de Tensão
P_{med}	Potência média no Transformador
P_{Ncu}	Perdas de Potência no Cobre do Transformador em Condição Nominal de Carga
P_{nom}	Potência Nominal do Transformador
r	Resistência por Unidade de Comprimento
R\$	Reais
R^2	Parâmetro de Homogeneidade
R^{TRECHO}	Resistência do Trecho classificado como Tronco
RT^{SDMT}	Resistência Total do Tronco em SDMT
SR^2	Variação Incremental de R^2
SSE	Soma Incremental dos Quadrados
V	Volt
$V_{\text{F}}^{1\text{f}}$	Tensão de Fase de Unidades Consumidoras Alimentadas em 1 Fases e 2 Fios
$V_{\text{F}}^{2\text{f}}$	Tensão de Fase de Unidades Consumidoras Alimentadas em 2 Fases e 3 Fios

V_F^{2f1}	Tensão de Fase de Unidades Consumidoras Alimentadas em 1 Fases e 3 Fios
V_F^{3f}	Tensão de Fase de Unidades Consumidoras Alimentadas em 3 Fases e 4 Fios
V_L	Tensão Média no Tronco
W	Watt

SUMÁRIO

1	INTRODUÇÃO	18
1.1	CONTEXTUALIZAÇÃO	18
1.1.1	Sistema Elétrico de Potência	18
1.1.2	Perdas no Sistema Elétrico de Potência.....	21
1.2	MOTIVAÇÕES	22
1.3	OBJETIVOS	23
1.4	ESTRUTURA DO TRABALHO	24
2	FUNDAMENTAÇÃO TEÓRICA.....	25
2.1	CÁLCULO DE PERDAS TÉCNICAS PELA ANEEL	25
2.1.1	Premissas de Cálculo	25
2.1.2	Metodologia para o Cálculo das Perdas Técnicas de Potência.....	26
2.2	CLUSTERIZAÇÃO	31
2.2.1	Conceituação	31
2.2.2	Medidas de Similaridade	32
2.2.3	Métodos de Clusterização.....	32
2.2.4	Algoritmos de Clusterização	34
2.2.5	Homogeneidade e Heterogeneidade	35
2.3	ANÁLISE DE REGRESSÃO LINEAR MÚLTIPLA	36
2.3.1	Normalidade Multivariada.....	37
2.3.2	Relação Linear Entre Variáveis	38
2.3.3	Regressão Linear <i>Stepwise</i>	39
3	MATERIAIS E MÉTODOS.....	41
3.1	DADOS UTILIZADOS.....	41
3.1.1	Levantamento dos Dados.....	41
3.2	CLUSTERIZAÇÃO	45
3.3	ANÁLISE DE REGRESSÃO	45
4	APRESENTAÇÃO E ANÁLISE DOS RESULTADOS	48

4.1	NÍVEIS DAS PERDAS TÉCNICAS NAS CONCESSIONÁRIAS DE DEISTRIBUIÇÃO BRASILEIRAS	48
4.2	CLUSTERIZAÇÃO	50
4.2.1	Definição do algoritmo a ser aplicado	51
4.2.2	Clusterização das distribuidoras com base nos percentuais das perdas técnicas em SDMT e SDBT	53
4.3	INVESTIGAÇÃO DOS ATRIBUTOS MAIS EXPLICATIVOS DAS PERDAS TÉCNICAS EM NÍVEIS DE MÉDIA E BAIXA TENSÃO	57
4.3.1	Análise da distribuição das variáveis utilizadas para estudo das concessionárias brasileiras.....	57
4.3.2	Análise de correlações	58
4.3.3	Avaliação da multicolinearidade das variáveis independentes.....	62
4.3.4	Seleção de variáveis por meio de análise de regressão <i>stepwise</i>	63
5	CONCLUSÕES.....	69
5.1	CONTRIBUIÇÕES DO TRABALHO.....	69
5.2	SUGESTÃO DE TRABALHOS FUTUROS.....	70
6	REFERÊNCIAS BIBLIOGRÁFICAS	72
	APÊNDICES	75
	APÊNDICE A Perdas na Distribuição	76

1 INTRODUÇÃO

1.1 CONTEXTUALIZAÇÃO

1.1.1 Sistema Elétrico de Potência

A energia elétrica é um importante elemento no processo de modernização, impulsionando a industrialização, alterando a estrutura urbana e refletindo a cultura das sociedades. No Brasil, a eletricidade é responsável por grande parte da energia consumida, representando cerca de 17,5% do consumo energético final do país (EPE, 2017). O consumo de energia elétrica fica atrás apenas do óleo diesel, conforme ilustra a Figura 1.1.

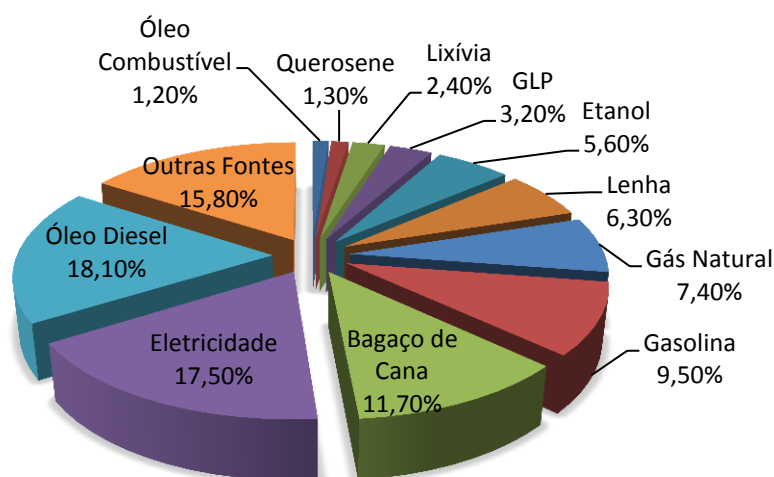


Figura 1.1 - Consumo final de energia por fonte.

(Fonte: EPE, 2017)

Os sistemas elétricos de potência são normalmente representados pela divisão clássica em geração, transmissão e distribuição da energia elétrica. A geração é a responsável por produzir a energia requisitada pelos consumidores. A transmissão faz o transporte de grandes blocos de energia elétrica em redes com tensão superior a 230 quilovolts (kV). A distribuição, por sua vez, fica responsável pelo transporte de montantes menores da energia elétrica, e suas redes operam com tensão inferior a 230 kV.

Atualmente o Brasil possui um total 4.744 unidades geradoras em operação, totalizando 155.429.704 kW de potência instalada. Desses empreendimentos, 219 são usinas hidrelétricas (UHE), 2.937 usinas termelétricas (UTE), 476 usinas eólicas (EOL), 430 Pequenas Centrais Hidrelétricas (PCHs - usinas com potência instalada entre 3 e 30

megawatts (MW)), 2 usinas termonucleares (TNE), 619 Centrais Geradoras Hidrelétricas (CGHs - usinas com potência instalada igual ou inferior a 5MW) e 61 usinas fotovoltaicas (UFV). Ainda, está prevista para os próximos anos uma adição de 21.663.039 kW na capacidade de geração do País, proveniente dos 244 empreendimentos atualmente em construção e mais 549 em Empreendimentos com Construção não iniciada (ANEEL, 2017a). A Tabela 1.1 ilustra o montante de unidades geradoras no Brasil, bem como suas potências outorgada e fiscalizada para o ano de 2017.

Tabela 1.1 - Unidades Geradoras no Brasil e Capacidade de Geração.

(Fonte: ANEEL, 2017a)

Tipo	Quantidade	Potência Outorgada (kW)	Potência Fiscalizada (kW)	%
UHE	219	101.188.678,00	94.488.995,00	60,79
UTE	2937	42.751.632,00	41.372.466,00	26,62
EOL	476	11.707.239,00	11.668.243,00	7,51
PCH	430	4.969.491,00	4.960.209,00	3,19
UTN	2	1.990.000,00	1.990.000,00	1,28
CGH	619	562.257,00	564.713,00	0,36
UFV	61	387.013,00	385.078,00	0,25
Total	4744	163.556.310,00	155.429.704,00	100

O segmento de transmissão no Brasil é dividido em dois grandes blocos: o Sistema Interligado Nacional (SIN) e os Sistemas Isolados. O Operador Nacional do Sistema Elétrico (ONS) é responsável pelo planejamento e operação de ambos. O SIN abrange quase a totalidade do território brasileiro e é constituído por quatro subsistemas: Sul, Sudeste/Centro-Oeste, Nordeste e a maior parte da região Norte. Em 2016, a extensão total desse sistema para transmissão era de 134.765 km, com previsão de 154.748 km para 2019 (ONS, 2016). A Figura 1.2 mostra o mapa do SIN. Os Sistemas Isolados, por outro lado, concentram-se na região Norte, nos estados de Rondônia, Acre, Amazonas, Roraima, Amapá e Pará. Eles estão presentes também na ilha de Fernando de Noronha, em Pernambuco, e em algumas localidades de Mato Grosso. Esses sistemas atendem atualmente 246 localidades no Brasil, cujo consumo representa menos de 1% da carga total do País. A demanda por energia dessas regiões é suprida, principalmente, por térmicas a óleo diesel (ONS, 2017b).

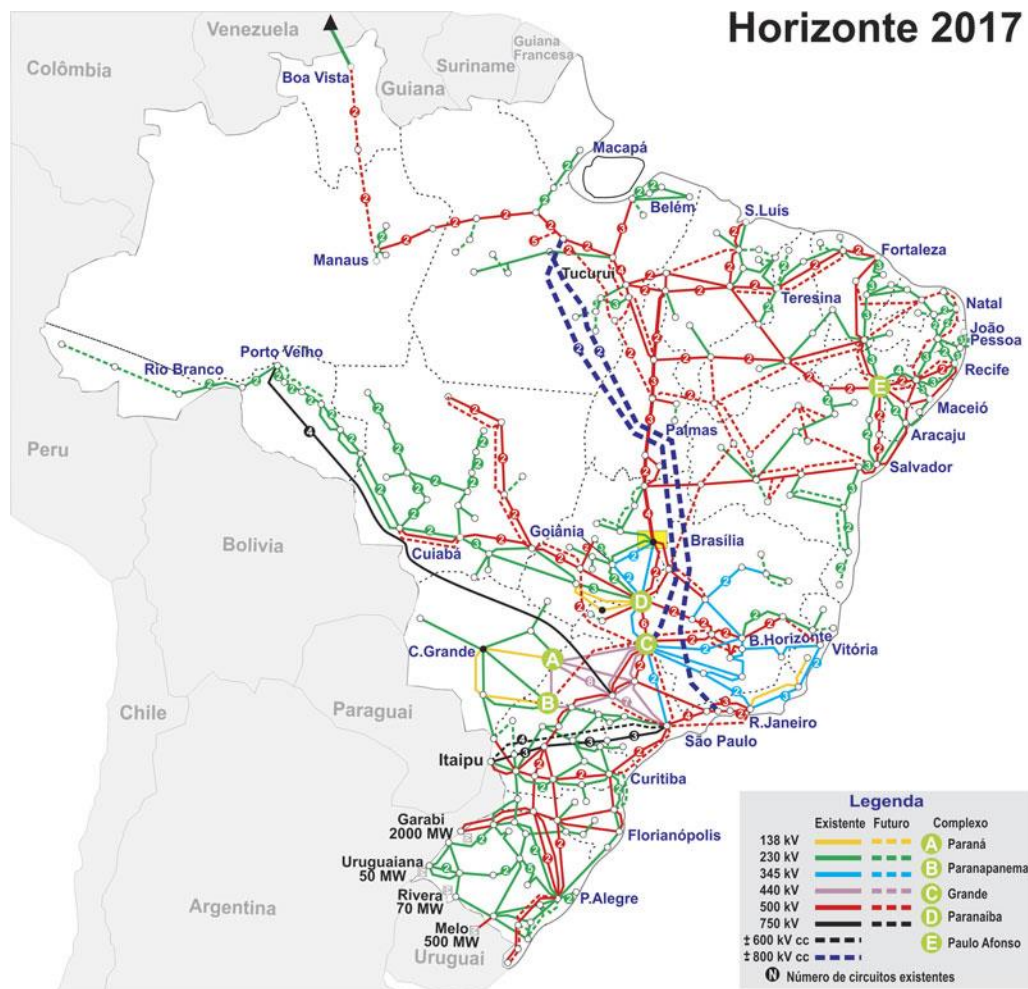


Figura 1.2 - Mapa do SIN 2017.

(Fonte: ONS, 2017a)

O mercado de distribuição de energia elétrica era formado por 101 empresas ao final do Terceiro Ciclo de Revisões Tarifárias Periódicas, sendo elas divididas em 63 concessionárias e 38 permissionárias (cooperativas de eletrificação rural para prestação do serviço público de distribuição). Atualmente, algumas distribuidoras foram agrupadas, resultando em 58 empresas de distribuição em atuação no país. As distribuidoras são responsáveis pelo acesso de quase 77 milhões de Unidades Consumidoras (UC) no Brasil. As concessionárias são responsáveis pelo fornecimento de energia para mais de 99% do mercado e do número de consumidores, representando quase a totalidade da prestação do serviço de distribuição no país (ANEEL, 2017b). A Figura 1.3 mostra o mapa das concessões de distribuição de energia elétrica no Brasil à época da realização dos cálculos para o Terceiro Ciclo de Revisões Tarifárias Periódicas.



Figura 1.3 - Mapa das concessões de distribuição de energia elétrica no Brasil.

(Fonte: ANEEL, 2017c)

1.1.2 Perdas no Sistema Elétrico de Potência

Os sistemas elétricos de potência possuem perdas em todas suas etapas: geração, transmissão e distribuição. As perdas no transporte da energia elétrica são usualmente classificadas em técnicas ou não técnicas. As primeiras são perdas devido ao processo físico do transporte (condução e transformação) da energia, enquanto que as demais perdas decorrem principalmente de furto (ligação clandestina, desvio direto da rede) ou fraude de energia (adulterações no medidor), erros de medição e de faturamento.

A Tabela 1.2 expõe os valores de perdas percentuais e sua evolução entre os anos 2000 e 2014 para diversos países. Apresentam-se também nessa tabela as médias de perdas para o mundo e para a América Latina. O percentual de perdas exposto considera as redes de transmissão e distribuição em cada país avaliado. Percebe-se que países desenvolvidos e mais ricos apresentam perdas inferiores às dos países em desenvolvimento e à média global. Nota-

se ainda que o Brasil apresentou em 2014 um percentual de perdas acima das médias da América Latina e do globo.

Tabela 1.2 - Percentuais de perdas na transmissão e distribuição para diversos países entre 2000 e 2014.

(Fonte: *World Energy Council*, 2017)

País	2000 (%)	2005 (%)	2010 (%)	2011 (%)	2012 (%)	2013 (%)	2014 (%)	2000 - 2014 (%/ano)
Luxemburgo	0,70	1,85	1,82	1,82	1,90	1,91	1,91	7,40
Islândia	4,18	4,48	4,17	2,99	2,82	2,10	2,10	-4,79
Alemanha	6,37	5,16	4,20	4,39	4,35	4,40	3,74	-3,73
Japão	4,67	4,83	4,64	4,88	4,61	4,75	4,75	0,12
Estados Unidos	6,00	6,60	6,28	6,27	6,56	6,19	6,15	0,18
Bolívia	10,22	10,30	11,77	11,25	11,58	9,25	9,63	-0,42
Colômbia	22,76	20,62	15,90	12,87	12,74	12,23	12,28	-4,31
Marrocos	7,49	6,34	10,29	10,58	10,92	13,74	13,74	4,43
Argentina	15,19	15,04	12,98	13,88	13,42	14,89	14,89	-0,14
Brasil	15,78	15,26	15,73	15,56	16,08	15,53	15,12	-0,30
Gana	18,39	24,09	25,71	24,60	22,68	22,56	23,58	1,79
Nepal	21,32	26,87	28,66	26,23	24,94	24,54	24,38	0,96
Paraguai	24,25	33,32	32,90	31,20	31,86	27,91	27,08	0,79
Mundo	9,41	9,42	8,73	8,64	8,70	8,61	8,63	-0,62
América Latina	16,36	16,53	15,67	15,30	15,33	15,04	14,85	-0,69

1.2 MOTIVAÇÕES

Como exposto na Tabela 1.2, a eficiência tocante às perdas na distribuição no Brasil encontra-se abaixo das médias mundial e latinoamericana. É importante pesquisar com cautela os fatores mais influentes para tal, dado que o custo dessas perdas é incorporado à tarifa paga pelo consumidor, segundo versam os Procedimentos de Regulação Tarifária (PRORET). Tanto as perdas na rede básica como as perdas na rede de distribuição são levadas em consideração nos processos tarifários que definem o valor final das tarifas de energia de cada empresa, representando um custo para o setor elétrico.

As perdas na rede básica são calculadas pela diferença entre as energias gerada e entregue nas redes de distribuição. Essas perdas são apuradas mensalmente pela Câmara de Comercialização de Energia Elétrica (CCEE) e o seu custo, que é definido anualmente nos processos tarifários, é rateado em 50% para geração e 50% para os consumidores (ANEEL, 2017d).

Por outro lado, os percentuais regulatórios de perdas técnicas e não técnicas na rede de distribuição são definidos pela ANEEL na Revisão Tarifária Periódica, que ocorre a cada 4 ou

5 anos. As perdas técnicas são calculadas conforme as regras definidas no Módulo 7 dos Procedimentos de Distribuição (PRODIST). A agência reguladora observa os níveis eficientes dessas perdas conforme as características da rede elétrica de cada concessionária. As perdas não técnicas, por sua vez, são apuradas pela diferença entre as perdas totais e as perdas técnicas. É importante frisar que as perdas totais são calculadas pela diferença entre a energia injetada na distribuição e a fornecida pela rede.

Segundo a agência, o nível médio de perda de distribuição no Brasil gera um custo anual superior a R\$ 10,5 bilhões no setor. Além disso, há discrepâncias entre as empresas, uma vez que os percentuais de perdas totais variam de 3% a 39% do mercado das mesmas, sendo as maiores divergências devidas à diferença no nível de perdas não técnicas entre as concessionárias de distribuição. Vale ressaltar, no entanto, que a variação nas perdas técnicas entre elas também é expressiva, indo de 3% a 14% da energia injetada total (ANEEL, 2016).

O Apêndice A traz em detalhes os valores de perdas técnicas e não técnicas calculadas no Terceiro Ciclo de Revisões Tarifárias Periódicas para as distribuidoras brasileiras. Nesse anexo, estão ausentes os dados relativos às empresas Companhia de Eletricidade do Amapá (CEA), Companhia Energética de Roraima (CERR) e Força e Luz Coronel Vivida Ltda (FORCEL) devido à carência de dados constatada pela ANEEL para a realização do cálculo de perdas para elas.

Dadas as discrepâncias existentes entre os níveis de eficiência tocante às perdas das concessionárias de distribuição no que tange perdas técnicas de energia, acredita-se necessário um estudo sobre as motivações de tais divergências. Sabe-se que o Brasil é um país com dimensões continentais e, portanto, ele apresenta áreas de distribuição com características bastante distintas umas das outras. Em decorrência desse fato, as redes elétricas que atendem cada uma das regiões brasileiras são também diversas e devem ser melhor avaliadas para uma determinação das particularidades que as fazem ter mais ou menos perdas técnicas no cenário nacional.

1.3 OBJETIVOS

O presente trabalho tem como objetivo apresentar uma análise das concessionárias brasileiras segundo seus níveis de perdas técnicas nos Sistemas de Distribuição de Média e Baixa Tensão (SDMT e SDBT). Objetiva-se também analisar as principais características das

redes elétricas brasileiras que influenciam os níveis de perdas técnicas na distribuição de energia no país.

Para tanto, são propostos i) a divisão das distribuidoras em grupos com níveis semelhantes de perdas técnicas em SDMT e SDBT e ii) posterior estudo das características comuns às empresas de um mesmo agrupamento, de forma a observar atributos que mais expliquem seus níveis de eficiência tocante a perdas técnicas na distribuição.

1.4 ESTRUTURA DO TRABALHO

Após esta introdução, o capítulo 2 aborda a metodologia de cálculo de perdas técnicas adotada pela ANEEL, que culmina nos valores de perdas técnicas em média tensão e baixa tensão (MT e BT) utilizados neste trabalho. Nesse capítulo também são apresentados aspectos dos processos de clusterização, bem como as técnicas de regressão multivariada necessárias para a avaliação dos parâmetros das redes de distribuição brasileiras.

No capítulo 3, exibem-se os materiais e métodos utilizados. Expõem-se as variáveis referentes às redes de distribuição brasileiras que possuem relevância para o resultado de perdas técnicas apresentado pelas distribuidoras nacionais. Complementarmente, descrevem-se os algoritmos avaliados para a clusterização das concessionárias e aborda-se a avaliação das variáveis dependentes e independente para a aplicação do método de regressão linear empregado neste estudo.

O capítulo 4 expõe e analisa os resultados das perdas técnicas nos níveis de média e baixa tensão para as 60 concessionárias de distribuição analisadas. Também se discorre sobre os resultados da clusterização executada com os três algoritmos disponíveis no software SAS® *Enterprise Guide* 7.1. Finalizando o capítulo, é exibida a análise das variáveis e posterior aplicação do método de regressão *stepwise* para determinação dos parâmetros mais explicativos de perdas técnicas em MT e BT para cada *cluster* formado no processo anterior.

O capítulo 5 apresenta as conclusões do trabalho assim como sugestões para trabalhos futuros.

2 FUNDAMENTAÇÃO TEÓRICA

2.1 CÁLCULO DE PERDAS TÉCNICAS PELA ANEEL

Os resultados das perdas técnicas nos Sistemas de Distribuição em Média e Baixa Tensão (SDMT e SDBT) utilizados neste trabalho foram retirados das notas técnicas referentes ao Terceiro Ciclo de Revisões Tarifárias Periódicas disponibilizadas no sítio eletrônico da Agência Nacional de Energia Elétrica (ANEEL). Os valores aqui empregados foram calculados pela Agência segundo a Revisão 3 do Módulo 7 dos Procedimentos de Distribuição de Energia Elétrica no Sistema Elétrico Nacional (PRODIST).

Esta seção visa esclarecer as premissas de cálculo adotadas pela ANEEL, bem como a metodologia empregada pela Agência para a mensuração das perdas técnicas nos sistemas de distribuição de média e baixa tensão.

2.1.1 Premissas de Cálculo

As premissas de cálculo da Revisão 3 do Módulo 7 incluem a adoção do valor de referência 0,92 para o fator de potência utilizado nos cálculos das perdas nos segmentos. Adicionalmente, a distribuição das cargas nas fases das redes SDMT é considerada equilibrada.

Outra consideração realizada por essa Revisão é a adição de 15% de perdas sobre o montante de perdas técnicas calculado para as redes do SDBT. Essa estimativa é feita devido ao posicionamento assimétrico do transformador em relação às tipologias de rede e ao desequilíbrio da carga.

Os níveis de tensão nominal de operação de cada concessionária são considerados para o cálculo das perdas técnicas. Além disso, é considerada uma temperatura constante e igual a 55°C dos condutores elétricos para a determinação das resistências ôhmicas dos mesmos.

As perdas nos transformadores são baseadas nos valores normatizados pela Associação Brasileira de Normas Técnicas (ABNT) para avaliação das perdas totais à carga nominal e a vazio. Na ausência destes, são adotados valores de projeto dos equipamentos informados pela distribuidora ou são utilizados valores típicos adotados pela ANEEL. Para este cálculo, considera-se a potência nominal de cada equipamento informado pelas empresas de distribuição.

Finalmente, são adicionados 5% de perdas sobre o montante de perdas técnicas totais calculadas devido às perdas produzidas por fugas de correntes em isoladores de para-raios e por efeito corona em conexões, sistemas supervisórios e em componentes da rede elétrica de distribuição, tais como relés fotoelétricos, capacitores, transformadores de corrente e de potencial.

Frisa-se ainda que o período de apuração das perdas técnicas por concessionária é, preferencialmente, anual.

2.1.2 Metodologia para o Cálculo das Perdas Técnicas de Potência

Nesta metodologia, as perdas de energia nas redes do Sistema de Distribuição em Alta Tensão (SDAT) são apuradas por dados dos sistemas de medição das próprias concessionárias e, portanto, elas não serão abordadas neste trabalho. Para o cálculo das perdas técnicas em SDMT e SDBT, por outro lado, são obtidas as perdas de potência e, posteriormente, são calculadas as perdas de energia para cada segmento de rede mediante a aplicação do Coeficiente de Perdas (CP). O CP representa a razão entre a perda média de potência e a perda de potência para a demanda média de cada segmento ou equipamento aqui considerados.

É importante salientar que a ANEEL adota a demanda média de potência calculada a partir das curvas de carga das concessionárias para o cálculo das perdas em SDMT e SDBT. As curvas de carga, por sua vez, são obtidas de dados do sistema de medição da distribuidora e da campanha de medição definida no Módulo 2 – Planejamento da Expansão do Sistema de Distribuição.

2.1.2.1 Perda de Potência e de Energia para o SDMT

Um modelo de regressão linear múltipla é utilizado para calcular as perdas de potência para a demanda média nas redes do SDMT. Para esta modelagem, faz-se necessário definir o comprimento médio do condutor tronco e ramal, bem como a resistência do condutor tronco e a corrente média no alimentador.

De acordo com a ANEEL, o condutor tronco é definido como o trecho de alimentador que parte da subestação até o ponto mais próximo desta onde a maior corrente a jusante é menor ou igual à maior corrente de qualquer ramal a montante. Os condutores ramais, por sua vez, são todos aqueles segmentos não classificados como tronco.

Ainda segundo a agência, o comprimento do tronco é, no máximo, 90% do comprimento total do alimentador, enquanto o ramal adquire um comprimento mínimo referente a 10% do comprimento total.

A resistência do condutor tronco (RT^{SDMT}), em ohm/km, é obtida pela média ponderada da resistência dos condutores classificados como tronco, conforme equação (1) (ANEEL, 2013).

$$RT^{SDMT} = \frac{\sum_{i=1}^N CT_i^{TRECHO} \times R_i^{TRECHO}}{CT^{SDMT}} \quad (1)$$

Onde N é o número de trechos classificados como tronco, CT^{TRECHO} e R^{TRECHO} são o comprimento (em km) e a resistência do trecho (em ohm/km) com seção constante classificado como tronco. Finalmente, CT^{SDMT} é o comprimento total desses trechos (em km).

Complementarmente, a corrente média utilizada na modelagem é obtida com base na energia fornecida pelo alimentador (em MWh) e na tensão nominal de linha do mesmo (V_L), em volt, como é mostrado na equação (2) (ANEEL, 2013).

$$I^{SDMT} = \frac{E^{SDMT}}{\sqrt{3} \times V_L \times \cos\varphi \times 8,76} \quad (2)$$

Onde I^{SDMT} é a corrente média no alimentador (em A), E^{SDMT} é a energia fornecida pelo alimentador (em MWh/ano) e $\cos\varphi$ é o fator de potência do mesmo, estabelecido em 0,92.

O cálculo da perda de potência para a demanda média de cada alimentador em SDMT (ΔP_{SDMT}) é dada em MW pela equação (3) (ANEEL, 2013).

$$\Delta P_{SDMT} = 0,95 \times e^{\alpha} \times 10^{-3} \quad (3)$$

Onde α é obtido pela equação (4) (ANEEL, 2013).

$$(4)$$

$$\alpha = (-6,34 + 1,82 \times \ln(I^{SDMT}) + 0,77 \times \ln(CT^{SDMT}) + 0,39 \times \ln(RT^{SDMT}) + 0,16 \times \ln(CR^{SDMT}))$$

Finalmente, para o cálculo das perdas de energia em SDMT (ΔE_{SDMT}), faz-se necessária a utilização do CP_{SDMT} , dado que a perda de energia está relacionada ao carregamento da rede, enquanto o resultado de perdas de potência leva em consideração somente a condição de carga média. A apuração da perda de energia nas N redes do SDMT é realizada conforme apresentado na equação (5) (ANEEL, 2013).

$$\Delta E_{SDMT} = \sum_{i=1}^N \Delta P_{SDMT_i} \times CP_{SDMT} \times \Delta T \quad (5)$$

Onde as perdas de energia são dadas em MWh e ΔT é o período de tempo analisado, em horas.

2.1.2.2 Perda de Potência e Energia para o SDBT

As perdas técnicas em sistemas de baixa tensão, por outro lado, são apuradas com base em redes típicas. Existem cinco tipologias de rede segundo as quais cada rede em SDBT deve ser classificada. A Agência define a regra exposta na Tabela 2.1 para tal classificação, que considera distribuição de carga uniforme e modelo de carga constante em relação à tensão. Nessa tabela, L_{circ} é o comprimento total do circuito em SDBT em metros.

Tabela 2.1 - Tipologias de rede em SDBT definidas pela ANEEL.

	Tipologia 1	Tipologia 2	Tipologia 3	Tipologia 4	Tipologia 5
Regra	$L_{circ} \leq 100$	$100 < L_{circ} \leq 200$	$200 < L_{circ} \leq 350$	$350 < L_{circ} \leq 500$	$L_{circ} > 500$

Há também a classificação em cinco tipologias alternativas, que variam em 2, 4, 8, 16 ou 24 trechos elementares e podem ser informadas pela concessionária.

Cada circuito em SDBT deve estar associado a um transformador de distribuição, e as resistências do tronco e do ramal são definidas, respectivamente, como a menor e a segunda menor resistências (em ohm/km) dentre os condutores do circuito analisado.

A perda de potência para a demanda média de um trecho elementar em SDBT é calculada segundo a equação (6) (ANEEL, 2013).

$$\Delta P_{SDBT} = r \times l \left(\frac{i^2 \times l^2}{3} + i \times I_j \times l + I_j^2 \right) \times 10^{-6} \quad (6)$$

Onde ΔP_{SDBT} é a perda de potência (em MW) para a demanda média no trecho elementar do circuito SDBT, l é o comprimento elementar (em km) dado pelo comprimento total do circuito dividido pelo número de trechos elementares referente à respectiva tipologia, r é a resistência por unidade de comprimento (em ohm/km), I_j é a corrente total (em A) a jusante do trecho elementar e i é a densidade de corrente (em A/km) dada pela razão entre a corrente máxima do circuito e seu comprimento total.

Analogamente ao que é realizado para a obtenção das perdas de energia no SDMT, o cálculo dessas perdas no SDBT (ΔE_{SDBT}) é dado em MWh pela equação (7) (ANEEL, 2013).

$$\Delta E_{SDMT} = \sum_{i=1}^N \Delta P_{SDBT_i} \times CP_{SDBT} \times \Delta T \quad (7)$$

Onde CP_{SDBT} é o Coeficiente de Perdas no SDBT e ΔT é o período de tempo analisado, em horas.

2.1.2.3 Perda de Potência e Energia para as Unidades Transformadoras

A equação (8) exhibe o cálculo realizado para o montante de perda de potência para a demanda média em transformadores (ΔP_{TR}) em MW (ANEEL, 2013).

$$\Delta P_{TR} = (\Delta p_{fe}^t + \Delta p_{cu}^t) \times 10^{-3} \quad (8)$$

Onde Δp_{fe}^t é a perda (em kW) no ferro ou a vazão do trafo t e Δp_{cu}^t é a perda de potência (em kW) para a demanda média no cobre do mesmo.

A perda de potência para a demanda média no cobre do transformador, por sua vez, é calculada pela equação (9) (ANEEL, 2013).

$$\Delta p_{cu} = \left(\frac{P_{med}}{P_{nom} \times \cos\varphi} \right)^2 \times P_{Ncu} \quad (9)$$

Onde P_{med} é potência média (em kW) no transformador obtida pela razão entre a energia consumida pelos consumidores ligados ao transformador e o tempo, P_{nom} é a potência nominal do transformador (em kVA), $\cos\varphi$ é o fator de potência igual a 0,92 e P_{Ncu} representa as perdas no cobre do transformador (em kW) na condição nominal de carga, sendo obtida pela diferença entre as perdas totais e a perda a vazão do trafo.

As perdas de energia para as NT unidades transformadoras são obtidas, em MWh, pela soma das perdas de energia a vazio pelo período de tempo analisado (ΔT) com as perdas de energia ocorridas no cobre, calculadas pela multiplicação das perdas de potência para a demanda média no cobre pelo Coeficiente de Perdas do trafo (CP_T) e pelo período de tempo analisado (ΔT), conforme definido na expressão (10) (ANEEL, 2013).

$$\Delta E_T = \Delta T \times \sum_{t=1}^{NT} (\Delta p_{fe}^t + \Delta p_{cu}^t \times CP_T) \quad (10)$$

2.1.2.4 Perda de Potência e de Energia em Ramais de Ligação e Medidores

A perda de potência para demanda média em ramais de ligação (ΔP_R) é calculada tendo-se como base a resistência média (R) do condutor. São levados em consideração ainda a corrente média de fase (I_m^f), o fator de diversidade (F_d) fixado em 0,7 e o número de unidades consumidoras alimentadas em 3 fases e 4 fios (N_{UC}^3), em 2 fases e 3 fios (N_{UC}^2), em 1 fase e três fios ($N_{UC}^{2'}$) e em 1 fase e 2 fios (N_{UC}^1). A perda total, portanto, é dada pela equação (11) (ANEEL, 2013).

$$\Delta P_R = R \times \left(\frac{I_m^f}{F_d} \right)^2 \times (3 \times N_{UC}^3 + 3 \times N_{UC}^2 + 2 \times N_{UC}^{2'} + 2 \times N_{UC}^1) \times 10^{-6} \quad (11)$$

O resultado de ΔP_R é dado em MW, enquanto as unidades de R e I_m^f estão, respectivamente, em ohm e em ampère. Na modelagem adotada pela Agência, a corrente média de fase é calculada segundo a equação (12).

$$I_m^f = \frac{E_{forn}^B \times 10^6}{\cos \varphi \times (3 \times N_{UC}^3 \times V_F^{3f} + 2 \times N_{UC}^2 \times V_F^{2f} + 2 \times N_{UC}^{2'} \times V_F^{2f1} + N_{UC}^1 \times V_F^{1f}) \times \Delta T} \quad (12)$$

Onde V_F^{3f} , V_F^{2f} , V_F^{2f1} e V_F^{1f} correspondem às tensões de fase (em V) das unidades consumidoras alimentadas em 3 fases e 4 fios, em 2 fases e 3 fios, em 1 fase e 3 fios e em 1 fase e 2 fios nessa ordem. Ainda, $\cos \varphi$ é o fator de potência de referência adotado em 0,92.

As perdas nas bobinas de tensão dos medidores (ΔP_M) são computadas, em W, para as unidades consumidoras do grupo B. A ANEEL considera 1,0 W de perda por circuito de tensão (P_C) para medidores eletromecânicos e 0,5 W de P_C para medidores eletrônicos. Essas perdas devem ser multiplicadas pelo número de medidores do parque de medição da distribuidora para consumidores do SDBT, conforme a equação (13) (ANEEL, 2013).

$$\Delta P_M = P_C \times (3 \times N_{UC}^3 + 3 \times N_{UC}^2 + 2 \times N_{UC}' + 2 \times N_{UC}^1) \times 10^{-6} \quad (13)$$

Finalmente, as perdas de energia para os ramais de ligação e medidores (ΔE_{RM}) é calculada, em MWh, considerando-se o Coeficiente de Perdas nos ramais (CP_R) e o tempo analisado (ΔT) de acordo com a equação (14) (ANEEL, 2013).

$$\Delta E_{RM} = (\Delta P_R \times CP_R + \Delta P_M) \times \Delta T \quad (14)$$

2.2 CLUSTERIZAÇÃO

2.2.1 Conceituação

Clusterização, ou agrupamento, pode ser entendida como uma análise explanatória de dados que caracterizam diferentes objetos. Entender os dados e as informações presentes no mundo atual se mostra fundamental, uma vez que estes fornecem base para análises, raciocínios, decisões e, finalmente, para a compreensão de todos os tipos de objetos e fenômenos (Xu e Wunsch, 2009).

Segundo Hair et al. (2007), a clusterização pode ser definida como um grupo de técnicas multivariadas cuja finalidade primária é agregar objetos com base em características que eles possuem. Os agrupamentos formados devem ser tais que aquelas observações alocadas no mesmo grupo são mais parecidas umas com as outras do que com as de outros grupos. Isto é, objetiva-se maximizar a homogeneidade dentro de grupos, ao mesmo tempo em que se maximiza a heterogeneidade entre eles (Hair et al., 2007).

Dentro desse conceito, um *cluster*, ou grupo, pode ser descrito como um agregado de pontos no espaço de teste, de modo que a distância entre dois pontos de um determinado *cluster* é menor do que a distância entre qualquer ponto dentro desse *cluster* e qualquer ponto que não esteja nele (Everitt et al., 1980). Algoritmos de clusterização agrupam os objetos em análise em uma determinada quantidade de *clusters*.

Nos processos de agrupamento, os objetos são caracterizados por seus atributos. A partir deles, com base em uma técnica de cálculo de proximidade, obtém-se a distância entre os elementos e, assim, definem-se similaridades e dissimilaridades.

2.2.2 Medidas de Similaridade

A proximidade entre os objetos, ou medida de similaridade, pode ser calculada segundo uma extensa gama de medidas de distâncias. A mais comum é a distância Euclidiana, cuja matriz de distâncias é dada conforme a equação (15) (Timm, 2002):

$$D_{rs}^2 = \sum_{i=1}^n [(y_{ri} - y_{si})' x (y_{ri} - y_{si})] = \sum_{i=1}^n \| y_{ri} - y_{si} \|^2 \quad (15)$$

Onde y_r e y_s são dois vetores de objetos analisados, com n atributos característicos cada. Como uma mudança nas unidades de medida pode fazer com que uma variável domine o *ranking* das distâncias, a matriz de distância Euclidiana é mais eficaz para variáveis métricas (Timm, 2002).

Há ainda outras medidas de distâncias que podem ser utilizadas, entre elas as distâncias *City Block* e Mahalanobis (Timm, 2002). Enquanto as distâncias de *City Block* tendem a ter baixa sensibilidade para *outliers* (dados discrepantes), as de Mahalanobis, independentemente de as variáveis serem ou não proporcionais, podem mascarar *clusters*, dado que as distâncias dentro de um *cluster* tendem a se tornar maiores do que o intervalo entre eles (Hartigan, 1975).

2.2.3 Métodos de Clusterização

Existem diversos métodos de clusterização disponíveis na literatura, genericamente classificados em particionais ou hierárquicos (Xu e Wunsch, 2009).

Métodos particionais dividem diretamente os objetos em um número pré-estabelecido de *clusters*, sem o emprego de uma estrutura hierárquica (Xu e Wunsch, 2009).

Métodos hierárquicos, por outro lado, envolvem a construção de uma estrutura do tipo árvore, também conhecida como dendrograma, podendo esta ser aglomerativa ou divisiva. Nos métodos aglomerativos, cada objeto é inicialmente considerado como um *cluster*. À medida que se avança nesse método, os dois agrupamentos mais próximos são combinados em um novo *cluster*, reduzindo, assim, o número de grupos em uma unidade a cada passo. Ao final da clusterização, todos os agrupamentos são reunidos em um único *cluster*. O método hierárquico divisivo, por outro lado, parte de um único agrupamento com todas as observações analisadas para a formação de grupos com somente um objeto em cada (Hair et

al., 2007). O número ótimo de *clusters* a serem formados é definido posteriormente, baseando-se na análise da homogeneidade e heterogeneidade dos agrupamentos realizados.

A Figura 2.1 exibe um exemplo de dendrograma de uma clusterização hierárquica com 6 objetos.

Da análise da Figura 2.1, nota-se que o dendrograma apresenta cada etapa do processo de cluterização e os grupos formados em cada uma. Percebe-se que as linhas verticais indicam os *clusters* formados e as linhas horizontais apontam os elementos pertencentes a cada grupo.

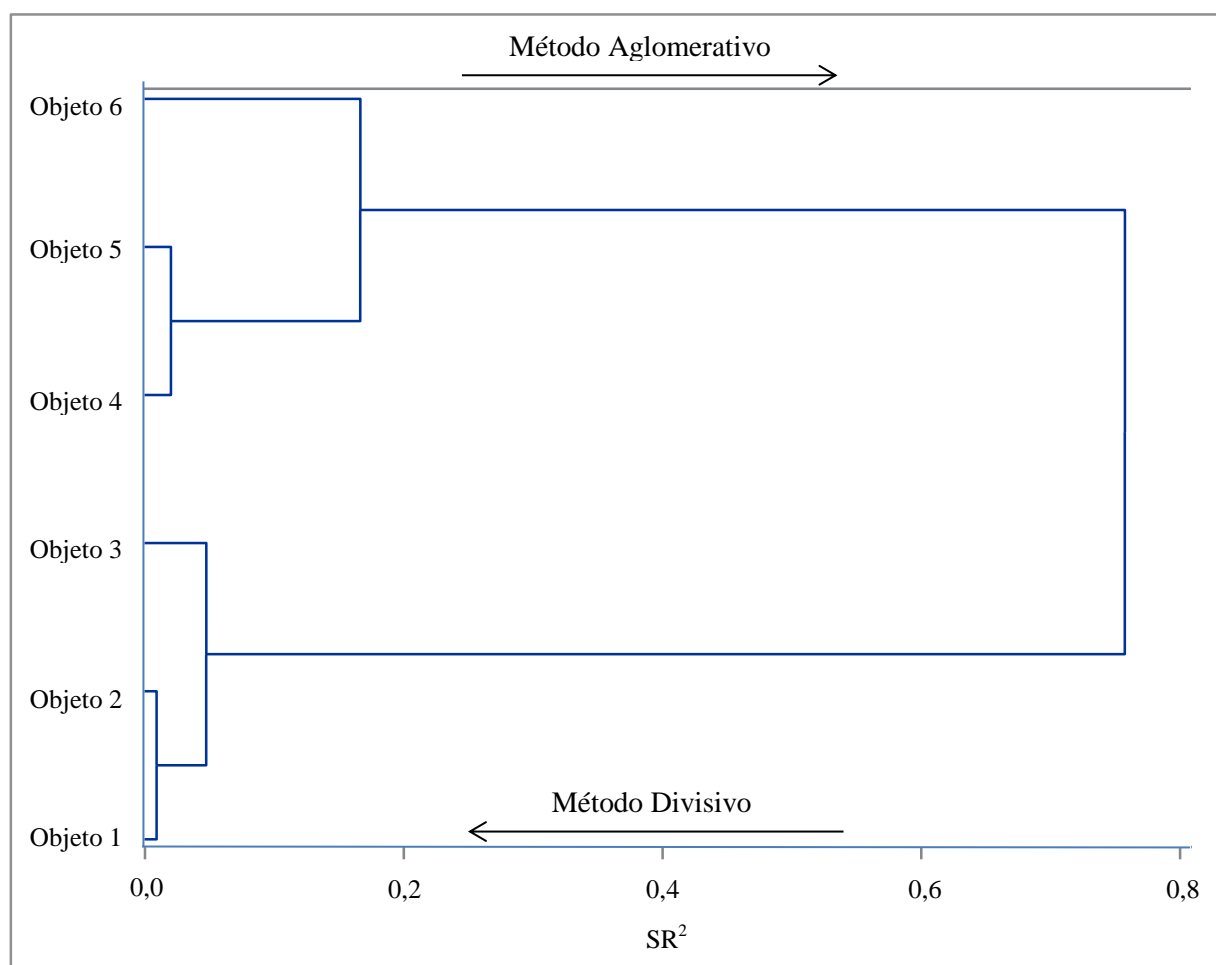


Figura 2.1 - Exemplo de dendrograma para clusterização hierárquica.

Considerando-se a clusterização aglomerativa, o primeiro agrupamento é constituído pelas observações 1 e 2, o segundo abrange a 4 e a 5, o terceiro aglutina o *cluster* que contém os objetos 1 e 2 com o objeto 3 e o quarto engloba o segundo grupo formado e o objeto 6. Por fim, ocorre a junção de todos eles em um único grupo.

No processo ilustrado na Figura 2.1 ao se desejar 3 grupos, por exemplo, obtém-se um *cluster* com os objetos 1, 2 e 3, um *cluster* com os objetos 4 e 5 e um *cluster* constituído pelo objeto 6, aqui classificado individualmente. Para a formação de 2 grupos, alternativamente, os elementos 4 e 5 são juntados com o 6.

2.2.4 Algoritmos de Clusterização

Algoritmos clássicos do método hierárquico aglomerativo são a ligação individual, ligação completa, média das distâncias, distância dos centroides e *Ward* (Hair et al., 2007). Neste trabalho, serão avaliadas apenas as clusterizações realizadas com base nos três últimos algoritmos e, por essa razão, tratam-se das únicas abordagens descritas a seguir.

2.2.4.1 Algoritmo da Média das Distâncias

O algoritmo de média das distâncias calcula a distância entre dois *clusters* segundo a equação (16) (Timm, 2002).

$$d_{(R)(S)} = \frac{\sum_r \sum_s d_{rs}}{n_R n_S} \quad (16)$$

Onde r pertence ao *cluster* R , s pertence ao *cluster* S , d_{rs} representa a medida de proximidade entre r e s e n_R e n_S representam o número de objetos em cada *cluster*.

2.2.4.2 Algoritmo da Distância dos Centroides

Segundo Timm (2002), partindo da matriz de proximidades, o algoritmo da distância dos centroides executa a combinação entre agrupamentos usando a média ponderada dos dois *clusters* mais similares, formando um novo *cluster* T .

Considerando-se dois *clusters* R e S , seus centroides são apresentados na equação (17) (Timm, 2002).

$$\bar{y}_R = \frac{\sum_r y_r}{n_R} \quad e \quad \bar{y}_S = \frac{\sum_s y_s}{n_S} \quad (17)$$

Onde \bar{y}_R e \bar{y}_S representam os centroides dos *clusters* R e S , nessa ordem. Ainda, y_r denota os objetos do *cluster* R e y_s corresponde aos objetos do *cluster* S .

A distância euclidiana entre os *clusters* R e S é dada pela equação (18).

$$d_{rs}^2 = \| \bar{y}_r - \bar{y}_s \|^2 \quad (18)$$

O centroide do novo *cluster* T, \bar{y}_t , formado pela agregação de R e S é apresentado em (19) (Timm, 2002).

$$\bar{y}_t = \frac{(n_R \times \bar{y}_r + n_S \times \bar{y}_s)}{n_R + n_S} \quad (19)$$

2.2.4.3 Algoritmo de Ward

No método de *Ward*, também conhecido como método de mínima variância ou soma incremental dos quadrados, o objetivo é minimizar o crescimento da soma dos erros quadráticos nos *clusters*, sendo essa definida para um único grupo com n objetos conforme a equação (20) (Timm, 2002).

$$SSE = \sum_{i=1}^n (y_i - \bar{y})' x (y_i - \bar{y}) = \sum_{i=1}^n \| y_i - \bar{y} \|^2 \quad (20)$$

Onde y_i representa os elementos de um dado *cluster* e \bar{y} é o seu centroide.

Combinando-se os *clusters* R e S de modo a formar um novo *cluster* T, a soma dos erros quadráticos do último grupo é dada pela equação (21) (Timm, 2002).

$$SSE_t = \sum_t \| y_t - \bar{y}_t \|^2, \quad t \in T \quad (21)$$

Portanto, o acréscimo incremental da junção de R e S é representado pela equação (22) (Timm, 2002).

$$SSE_t - (SSE_r + SSE_s) \quad (22)$$

Finalmente, para a aglutinação dos *clusters*, o algoritmo de *Ward* procura o menor valor do incremento da junção de dois *clusters* distintos.

2.2.5 Homogeneidade e Heterogeneidade

Os parâmetros de homogeneidade e heterogeneidade são importantes para a avaliação dos algoritmos e para a determinação do número ótimo de *clusters* a serem formados com uma determinada quantia de objetos.

Em análise de clusterização, o parâmetro de homogeneidade é descrito por R^2 , cujo valor para k *clusters* é calculado pela expressão (23) (Timm, 2002).

$$R_k^2 = \frac{SSE - \sum_k SSE_k}{SSE} \quad (23)$$

Onde SSE é a soma total dos erros quadráticos nos *clusters* para n objetos conforme descrito em (20). Para n *clusters*, cada SSE_k é nulo, implicando em R^2 igual a 1. Nesse cenário, tem-se o máximo de homogeneidade dentro dos *clusters* e de heterogeneidade entre eles. No entanto, à medida que o número de agrupamentos decresce de n para 1, R^2 também diminui.

A variação em R^2 devido à junção de dois *clusters* é representada por SR^2 , conforme a equação (24) (Timm, 2002).

$$SR^2 = R_k^2 - R_{k-1}^2 \quad (24)$$

Onde k é o número de clusters formados com os n objetos em análise. Quanto maior o incremento SR^2 , maior é a perda de homogeneidade dentro dos agrupamentos.

2.3 ANÁLISE DE REGRESSÃO LINEAR MÚLTIPLA

Análise de regressão linear múltipla é uma técnica estatística que pode ser utilizada para examinar a relação entre uma única variável dependente e várias variáveis independentes (Hair et al., 2007).

As aplicações da regressão múltipla podem ser classificadas em dois grandes grupos de problemas, a saber: previsão e explicação. Neste trabalho, a análise de regressão é utilizada com o objetivo explicativo. Aqui, busca-se compreender quais são os fatores que mais influenciam os resultados das perdas técnicas nos sistemas de distribuição em média e baixa tensão nas redes brasileiras.

A regressão múltipla fornece uma interpretação objetiva da importância relativa de cada variável independente na previsão da medida dependente. Nesta análise, é possível avaliar simultaneamente as relações entre cada atributo independente e a variável dependente, determinando-se a importância relativa de cada variável independente no modelo (Hair et al., 2007).

O modelo de regressão linear pode ser representado conforme exposto na equação (25) (Devore, 2006):

$$y = X.\beta + \epsilon \quad (25)$$

Onde X é a matriz com as variáveis independentes, β é o vetor de coeficientes e ϵ é o erro ou desvio aleatório.

Uma boa estimativa de y será uma reta que se ajuste bem aos dados observados. O ajuste, nesse caso, é normalmente avaliado com uma métrica de erros quadráticos, ou quadrados mínimos. Uma reta oferece um bom ajuste aos dados se os desvios dos pontos no espaço em relação à reta são pequenos. A medida de ajuste é a soma dos quadrados desses desvios. A reta de melhor ajuste é, portanto, a reta que tem a menor soma possível de desvios quadráticos (Devore 2006).

A aplicação da regressão linear pressupõe que os dados independentes e a variável dependente atendem a algumas suposições estatísticas: i) normalidade multivariada e ii) relação linear entre as variáveis (Timm, 2002). Ambos os tópicos serão abordados em maior detalhe nas seções seguintes.

De acordo com Hair et al. (2007), existem três abordagens de seleção de modelos de regressão linear mais utilizadas na literatura: i) método confirmatório, ii) método de busca sequencial e iii) método combinatório. Neste estudo, é utilizado o segundo método: modelo de regressão linear com busca sequencial *stepwise*, o qual é explicado na Seção 2.3.3.

2.3.1 Normalidade Multivariada

A suposição mais fundamental em análise multivariada é a normalidade, que se refere à forma da distribuição de dados para uma variável métrica individual. A distribuição normal é o padrão de referência para os métodos estatísticos (Hair et al., 2007).

Segundo Timm (2002), métodos estatísticos multivariados baseiam-se na suposição de normalidade multivariada, isto é, as variáveis individuais são normais em um sentido univariado, bem como suas combinações. É indispensável, portanto, a avaliação da distribuição para todas as variáveis presentes no modelo e possíveis transformações de dados quando a não-normalidade é detectada. Se a variação em relação à distribuição normal é

suficientemente grande, todos os testes estatísticos resultantes são inválidos (Hair et al., 2007).

As transformações de dados fornecem os principais meios para corrigir a não-normalidade da distribuição das variáveis. Segundo Hair et al. (2007), as transformações corretivas mais utilizadas são três, a saber: i) o logaritmo natural, ii) a raiz quadrada ou iii) o inverso da variável.

2.3.2 Relação Linear Entre Variáveis

2.3.2.1 Correlação de *Pearson*

A correlação de *Pearson* é uma medida paramétrica de associação linear para duas variáveis. Essa medida indica tanto a força quanto a direção de um relacionamento linear e pode variar de -1 a 1 (Mingoti, 2005). Correlação com módulo unitário revela uma associação perfeita entre duas variáveis, enquanto valores próximos a zero indicam baixa correlação entre elas.

A correlação de *Pearson* entre duas variáveis x e y é calculada segundo a equação (26) (Swinscow e Campbell, 1997).

$$r_{xy} = \frac{\sum_{i=1}^n ((x_i - \bar{x}) \cdot (y_i - \bar{y}))}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \cdot \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}} \quad (26)$$

Onde x_i e y_i são os valores medidos de ambas variáveis, e \bar{x} e \bar{y} são as médias aritméticas das mesmas.

É importante examinar a correlação entre as variáveis independentes e a dependente, visto que a relação linear entre elas é uma suposição estatística básica para a aplicação de métodos de regressão multivariados. A manutenção no modelo de variáveis pouco correlacionadas com a dependente fere essa suposição e diminui o poder preditivo da equação de regressão (Hair et al., 2007).

Além disso, a avaliação da correlação de *Pearson* das variáveis independentes entre si é importante para a análise de regressão múltipla, uma vez que uma correlação elevada entre esses atributos pode tornar algumas variáveis redundantes no esforço preditivo. Isso indica que, em um contexto multivariado, elas não são necessárias para produzir a previsão ótima

caso um outro conjunto de variáveis independentes que explique essas variáveis seja empregado.

2.3.2.2 Multicolinearidade

A habilidade de uma variável independente adicional de melhorar a previsão da variável dependente está relacionada não apenas à sua correlação com a variável dependente, mas também com as correlações da variável independente adicional com as variáveis independentes já incluídas na equação de regressão. A colinearidade é a associação medida entre duas variáveis independentes. A multicolinearidade, por sua vez, refere-se à correlação entre três ou mais variáveis independentes evidenciada quando uma é regressada em relação às demais (Hair et al., 2007).

O impacto da multicolinearidade é reduzir o poder preditivo de qualquer variável independente na medida em que ela é associada com as outras variáveis independentes. Portanto, segundo Hair et al. (2007), a previsão de uma variável dado um número de variáveis independentes é maximizada quando estas possuem baixa multicolinearidade entre si. Ao mesmo tempo, esses atributos devem possuir alta colinearidade com a variável que deseja ser explicada.

Uma forma de avaliar a multicolinearidade entre variáveis é por meio da estatística denominada fator de inflação de variância (VIF – *Variance Inflation Factor*). Um valor de VIF elevado indica que uma variável independente é bastante explicada pelas demais variáveis independentes presentes no modelo. Valores de VIF acima de 5 indicam uma presença substancial de multicolinearidade, resultando em uma correlação múltipla de 0,89.

2.3.3 Regressão Linear *Stepwise*

A estimação *stepwise* permite examinar a contribuição de cada variável independente para o modelo de regressão. Antes do desenvolvimento da equação, cada variável é considerada para inclusão no modelo. Aquela com maior contribuição ao resultado da variável dependente é acrescentada em um primeiro momento. Variáveis independentes são então selecionadas para inclusão com base em sua contribuição incremental sobre as variáveis já presentes na equação (Hair et al., 2007). Caso a adição de uma nova variável não resulte em uma contribuição significativa, o método *stepwise* pode eliminar a variável do modelo.

O passo a passo para a seleção de variáveis via regressão *stepwise* é descrito a seguir:

1. Inicia-se o modelo com a variável independente que é a mais fortemente correlacionada com a variável dependente, resultando em uma equação inicial do tipo $Y = b_0 + b_1.X_1$;
2. Os coeficientes de correlação parcial são examinados para encontrar uma variável independente adicional que explique a maior parte estatisticamente significativa do erro remanescente da primeira equação de regressão;
3. A equação de regressão é recalculada usando as duas variáveis independentes, e o valor parcial para a variável original, de modo a avaliar se a inserção da variável fez uma contribuição significativa para o resultado final. Se não for o caso, a variável é eliminada. Se a variável ainda fizer uma contribuição significativa, a equação passa a ser $Y = b_0 + b_1.X_1 + b_2.X_2$;
4. Esse procedimento é realizado novamente até que todas as variáveis independentes não-presentes no modelo sejam examinadas. O intuito dessa análise é determinar se alguma deveria ser incluída na equação para melhor estimação da variável dependente. Uma vez que uma nova variável é acrescentada, todas as variáveis independentes previamente adicionadas devem ser reavaliadas para julgar se elas devem ser mantidas.

Nota-se portanto que, neste método de regressão, somente uma variável é considerada para inclusão no modelo por vez e que há possibilidade de retirada de variáveis da equação no decorrer do processo de regressão.

3 MATERIAIS E MÉTODOS

Este capítulo tem como objetivo apresentar os materiais e os métodos utilizados i) para clusterização das concessionárias de distribuição de energia elétrica brasileiras de acordo com suas perdas técnicas nos níveis de média e baixa tensão, e ii) para a investigação dos principais geradores dessas perdas nas redes de distribuição do país em cada *cluster* formado.

Para tanto, inicialmente, expõem-se as variáveis referentes às redes de distribuição brasileiras que possuem relevância para o resultado de perdas técnicas apresentado pelas distribuidoras nacionais.

Na sequência, descrevem-se os algoritmos avaliados para a clusterização das concessionárias.

Fechando o capítulo, aborda-se a avaliação das variáveis dependentes e independente para a aplicação do método de regressão linear utilizado neste trabalho. Também é abordado o método de regressão *stepwise*, utilizado para a determinação dos parâmetros mais relevantes para as perdas em cada agrupamento formado no processo de clusterização.

3.1 DADOS UTILIZADOS

3.1.1 Levantamento dos Dados

Nesta pesquisa, o intenso trabalho de captação de informações passa pela consulta nos bancos de dados da ANEEL.

Os dados referentes às características das redes de distribuição utilizados neste estudo foram extraídos das Notas Técnicas relativas ao Terceiro Ciclo de Revisões Tarifária Periódica, realizado entre os anos de 2011 e 2014. É importante destacar que as Notas Técnicas têm caráter público e podem ser acessadas no endereço eletrônico da agência. Nelas estão disponibilizados os dados utilizados na metodologia e nos procedimentos para a apuração das perdas no sistema de distribuição de energia elétrica, regulamentados no Módulo 7 – Cálculo de Perdas na Distribuição do PRODIST, Revisão 3. Frisa-se também que foram escolhidos dados relativos à terceira revisão tarifária em detrimento da mais recente, realizada entre os anos de 2015 e 2018, devido ao fato de os resultados de perdas para a

última revisão ainda não se encontrarem finalizados para todas as distribuidoras brasileiras quando da confecção deste estudo. Ademais, os valores de perdas técnicas apurados no último ciclo de revisão até o momento de conclusão deste trabalho mostram-se próximos aos resultados obtidos no ciclo anterior para as concessionárias já analisadas pela Agência.

Foram extraídas da base de dados da ANEEL referente aos indicadores coletivos de continuidade DEC (Duração Equivalente de Interrupção por Unidade Consumidora) e FEC (Frequência Equivalente de Interrupção por Unidade Consumidora) as informações da área e do número de unidades consumidoras de cada distribuidora.

A partir dos dados obtidos das notas técnicas, foram criadas novas variáveis, de modo que a comparação entre as empresas seja realizada também com dados relativos à área, à extensão de rede de média e baixa tensão, e ao número de consumidores de cada empresa. Dessa maneira, é possível realizar uma comparação relativa entre as distribuidoras utilizando, além de dados brutos, dados de densidade.

A Tabela 3.1 apresenta a lista dos 44 dados das concessionárias utilizados neste trabalho bem como as siglas adotadas para cada variável. Destaca-se que a escolha das variáveis a serem empregadas neste estudo fundamentou-se na base de dados requerida pela ANEEL para o cálculo de perdas das empresas de distribuição conforme versa a Revisão 3 do Módulo 7 do PRODIST.

Destaca-se da análise da Tabela 3.1 que os dados referentes aos níveis A3a e A4 são agrupados e considerados somente como nível de média tensão (MT), visto que nem todas as distribuidoras possuem rede A3a em sua área de concessão. Esse fator dificultaria a análise para todas as empresas de forma homogênea caso fosse mantida a distinção entre os níveis supramencionados.

Ressalta-se que há uma grande carência de informações relativas à Companhia de Eletricidade do Amapá (CEA), à Companhia Energética de Roraima (CERR) e à empresa Força e Luz Coronel Vivida Ltda (FORCEL), impedindo a realização do estudo para essas distribuidoras. Por esse motivo, apesar de estarem atuando no Brasil 63 concessionárias de distribuição à época da realização do Terceiro Ciclo de Revisões Tarifárias Periódicas, executa-se a análise de clusterização e seleção de variáveis para as 60 empresas que são exibidas na Tabela 3.2.

Tabela 3.1 - Dados utilizados para a avaliação das concessionárias de distribuição brasileiras

Variável	Sigla	Unidade
Número de Transformador de Distribuição (TD)	N_TD	-
Potência Aparente Instalada de TD Total	S_InstTot_TD	MVA
Potência Instalada Média por TD	S_InstMed_TD	kVA
Densidade de Potência Nominal Instalada de TD por Área	Dens_SInstTD_Area	kVA/km ²
Densidade de Potência Instalada de TD por km de Rede MT e BT	Dens_SInstTD_Rede	kVA/km
Densidade de Potência Instalada de TD por UC	Dens_SInstTD_UC	kVA
Densidade de TD por Área	Dens_TD_Area	1/km ²
Densidade de TD por km de Rede MT e BT	Dens_TD_Rede	1/km
Densidade de TD por UC	Dens_TD_UC	TD/consumidor
Percentual de Potência Nominal Instalada em TD	Perc_SInst_TD	%
Percentual Médio de Utilização	UtilizMed_TD	%
Número de Reguladores	N_Regul	-
Potência Instalada Total de Reguladores	S_InstTot_Regul	MVA
Potência Instalada Média por Regulador	S_InstMed_Reg	kVA
Densidade de Reguladores por Área	Dens_Regul_Area	Regul/km ²
Densidade de Reguladores por km de rede MT	Dens_Regul_Rede	Regul/km
Densidade de Reguladores por UC	Dens_Regul_UC	Regul/consumidor
Percentual Médio de Utilização	UtilizMed_Regul	%
Número de circuitos MT	N_CircMT	-
Potência Média Total MT	P_MedTot_MT	MW
Densidade de Potência por km de rede MT	DensPot_MT_Rede	kW/km
Resistência Média dos Segmentos de Rede MT	ResMed_Segm_MT	Ohms/km
Comprimento Total Médio MT	CompTotMed_MT	km
Densidade de Rede MT por Área	DensRede_MT_Area	km/km ²
Percentual de Rede MT na Distribuidora	PercRede_MT	%
Número de circuitos BT	N_CircB	-
Potência Média Total BT	P_MedTot_B	MW
Densidade de Potência por km de rede BT	DensPot_B_Rede	kW/km
Resistência Média dos Segmentos de Rede BT	ResMed_Segm_B	Ohms/km
Resistência Média do Condutor Ramal BT	ResMed_Ramal_B	Ohms/km
Comprimento Total Médio	CompTotMed_B	km
Densidade de Rede B por Área	DensRed_B_Area	km/km ²
Percentual de Rede B na Distribuidora	PercRede_B	%
Energia Total Injetada	EnerTot_Inj	MWh
Energia Total Injetada em MT e BT	EnerTot_MTBT	MWh
Energia Total Fornecida	EnerTot_Forn	MWh
Densidade de Energia Fornecida Anual por Área	DensEnerg_Forn_Area	MWh/km ²
Densidade de Energia Fornecida Anual por km de Rede MT e BT	DensEnerg_Forn_Rede	MWh/km
Densidade de Energia Fornecida Anual por UC	DensEnerg_Forn_UC	MWh/UC
Área Total	AREAT	km ²
Número de Unidades Consumidoras	NUC	-
Densidade de Unidades Consumidoras por Área	Dens_UC_Area	1/km ²
Densidade de Unidades Consumidoras por km de Rede MT e BT	Dens_UC_Rede	1/km
Percentual de Perdas MT e BT	PercPerdas_MTBT	%

Tabela 3.2 - Concessionárias analisadas

CONCESSIONÁRIA	SIGLA
Aes Sul Distribuidora Gaúcha de Energia S/A	AES-SUL
Amazonas Distribuidora de Energia S/A	AME
Ampla Energia e Serviços S/A	AMPLA
Bandeirante Energia S/A	BANDEIRANTE
Boa Vista Energia S/A	BOA VISTA
Caiuá Distribuição de Energia S/A	CAIUÁ
Companhia Energética de Alagoas	CEAL
Ceb Distribuição S/A	CEB
Companhia Estadual de Distribuição de Energia Elétrica	CEEE
Celesc Distribuição S.A.	CELESC
Celg Distribuição S.A.	CELG
Centrais Elétricas do Pará S/A	CELPA
Companhia Energética de Pernambuco	CELPE
Companhia Energética do Maranhão	CEMAR
Cemig Distribuição S/A	CEMIG
Companhia Energética do Piauí	CEPISA
Centrais Elétricas de Rondônia S/A	CERON
Companhia Força e Luz do Oeste	CFLO
Companhia Hidroelétrica São Patrício	CHESP
Companhia Nacional de Energia Elétrica	CNEE
Companhia Campolarguense de Energia	COCEL
Companhia de Eletricidade do Estado da Bahia	COELBA
Companhia Energética do Ceará	COELCE
Cooperativa Aliança	COOPERALIANÇA
Copel Distribuição S/A	COPEL
Companhia Energética do Rio Grande do Norte	COSERN
Companhia Jaguari de Energia	CPFL JAGUARI
Companhia Leste Paulista de Energia	CPFL LESTEPAULISTA
Companhia Luz e Força de Mococa	CPFL MOCOCA
Companhia Paulista de Força e Luz	CPFL PAULISTA
Companhia Piratininga de Força e Luz	CPFL PIRATININGA
Companhia Luz e Força Santa Cruz	CPFL SANTACRUZ
Companhia Sul Paulista de Energia	CPFL SULPAULISTA
Departamento Municipal de Energia de Ijuí	DEMEI
Dme Distribuição S.A	DMED
Energisa Borborema Distribuidora de Energia S.A.	EBO
Empresa de Distribuição de Energia Vale Paranapanema	EDEVP
Empresa Elétrica Bragantina S/A	EEB
Empresa Força e Luz João Cesa Ltda	EFLJC
Empresa Força e Luz Urussanga Ltda	EFLUL
Elektro Eletricidade e Serviços S/A	ELEKTRO
Companhia de Eletricidade do Acre	ELETROACRE
Centrais Elétricas de Carazinho S/A	ELETROCAR
Eletropaulo Metropolitana Eletricidade de São Paulo S/A	ELETROPAULO
Empresa Luz e Força Santa Maria S/A	ELFSM
Energisa Minas Gerais - Distribuidora de Energia S.A.	EMG
Energisa Mato Grosso do Sul - Distribuidora de Energia	EMS
Energisa Mato Grosso - Distribuidora de Energia S.A	EMT
Energisa Nova Friburgo - Distribuidora de Energia S.A	ENF
Energisa Paraíba - Distribuidora de Energia S.A.	EPB
Espírito Santo Centrais Elétricas S/A	ESCELSA
Energisa Sergipe - Distribuidora de Energia S.A.	ESE
Energisa Tocantins - Distribuidora de Energia S.A.	ETO
Hidroelétrica Panambi S/A	HIDROPAN
Iguaçu Distribuidora de Energia Elétrica Ltda	IENERGIA
Light Serviços de Eletricidade S/A	LIGHT
Muxfeldt Marin & Cia. Ltda	MUXENERGIA
Rio Grande Energia S/A	RGE
Companhia Sul Sergipana de Eletricidade	SULGIPE
Usina Hidroelétrica Nova Palma Ltda	UHENPAL

3.2 CLUSTERIZAÇÃO

Neste trabalho, os processos de clusterização são executados por meio do *software* SAS[®] *Enterprise Guide* 7.1. Esse programa possui três algoritmos de clusterização hierárquica aglomerativa, a saber:

- i) Média das distâncias entre os objetos;
- ii) Distância dos centroides; e
- iii) *Ward*.

Neste estudo avalia-se qual das opções supracitadas apresenta o melhor resultado de agrupamento.

O atributo utilizado para o agrupamento das empresas é o percentual de perdas técnicas em níveis de média e baixa tensão adquirido das Notas Técnicas relativas ao Terceiro Ciclo de Revisões Tarifárias Periódicas, conforme exposto anteriormente.

Com base no resultado da clusterização, é possível inferir as variáveis mais explicativas de perdas técnicas em MT e BT para cada *cluster* por meio de uma análise de regressão.

3.3 ANÁLISE DE REGRESSÃO

Após a obtenção dos *clusters* baseados nas perdas técnicas em média e baixa tensão das concessionárias analisadas, é necessário realizar o estudo de quais variáveis são mais explicativas de perdas em cada agrupamento. Para tanto, realiza-se uma análise de regressão com os parâmetros apresentados na Tabela 3.1 para cada grupo.

Assim como realizado no trabalho de Vidal Rôla Almeida (2015), neste estudo foi escolhida a abordagem de regressão *stepwise* para a seleção de variáveis para aplicação em modelagem estatística.

A fim de se determinar quais parâmetros permanecerão no modelo para aplicação da regressão *stepwise*, são analisadas as suposições estatísticas da análise multivariada. Para tanto, verificam-se a normalidade univariada das variáveis independentes e dependente e a linearidade das mesmas. Neste trabalho, a variável dependente é a perda técnica em média e baixa tensão, e todos os demais parâmetros são considerados variáveis independentes.

A distribuição normal é o padrão de referência para todos os métodos estatísticos. Se a variação em relação à distribuição normal é suficientemente grande, todos os testes estatísticos resultantes são inválidos. Além disso, a linearidade é indispensável, porque as correlações representam apenas a associação linear entre variáveis; os efeitos não-lineares não são representados no valor de correlação (Hair et al, 2007).

Os passos realizados para a verificação de tais suposições e para uma filtragem inicial das variáveis que serão utilizadas no modelo de regressão *stepwise* são realizados na seguinte ordem:

- i) Avaliação da distribuição de cada uma das variáveis;
- ii) Avaliação da correlação de *Pearson* entre as variáveis independentes e a dependente;
- iii) Avaliação da correlação de *Pearson* das variáveis independentes entre si; e
- iv) Avaliação da multicolinearidade.

Os testes de normalidade são realizados para verificar se cada uma das variáveis (dependente e independentes) existentes no modelo possui uma distribuição normal. São utilizados os resultados dos três testes disponíveis no SAS *Enterprise Guide* 7.1[®]: Kolmogorov-Smirnov, Cramer-von Mises e Anderson-Darling. Além disso, faz-se necessária a transformação de algumas variáveis independentes para que sua distribuição passe a ser normal, atendendo à primeira suposição estatística. Primeiramente, é aplicado o logaritmo natural a essas variáveis, e os testes de normalidade são refeitos. Ao final desse passo, aplica-se a raiz quadrada nos parâmetros cuja distribuição ainda não é normal. Mais uma vez, os testes de normalidade são reavaliados. Em seguida, aplica-se a transformação do inverso das variáveis ainda não adequadas à distribuição normal e testa-se sua nova distribuição. Aquelas que, mesmo com essas transformações, não apresentarem uma distribuição normal, são retiradas do estudo.

As variáveis independentes são submetidas ao teste de correlação de *Pearson* para que uma filtragem inicial das mesmas possa ser realizada. A correlação é uma medida adequada para avaliar o grau de relacionamento linear entre duas variáveis quantitativas, com valores entre -1 e 1 (Mingoti, 2015). Os atributos com correlação entre si maior que 0,9 em módulo são considerados muito correlacionados e representativos de uma mesma grandeza.

A correlação de *Pearson* também é avaliada entre as variáveis independentes e a variável dependente. A avaliação da correlação neste caso visa basear a exclusão ou permanência das variáveis com correlação de *Pearson* entre si maior do que 0,9. Sendo assim, para cada par com essa característica, a variável com maior correlação com a variável dependente é mantida e a outra, excluída. As variáveis cujas correlações com o nível de perdas técnicas em MT e BT são muito baixas, isto é, abaixo de 0,01 em módulo, são retiradas do modelo.

Após a análise de correlação de *Pearson*, as variáveis independentes resultantes são avaliadas quanto à multicolinearidade por meio da estatística denominada fator de inflação de variância VIF. É importante frisar que o impacto da multicolinearidade é reduzir o poder preditivo de qualquer variável independente na medida em que ela é associada com outras variáveis independentes, como exposto no Capítulo 2.

Neste trabalho, antes da aplicação do modelo de regressão *stepwise*, é aplicado um modelo de regressão do tipo *Enter* para a avaliação da multicolinearidade dos parâmetros independentes. Neste passo, todas as variáveis são incluídas e os valores de VIF para cada variável é avaliado. Observando-se valores de VIF acima de 5, a variável com maior valor é excluída, e o modelo *Enter* é reestimado. Esse processo é realizado sucessivamente, retirando-se sempre a variável com o maior valor de VIF em cada passo, até que todas as variáveis apresentem VIF inferior a 5.

Ao término da análise da multicolinearidade, realiza-se a regressão *stepwise* para seleção das variáveis mais explicativas de perdas técnicas em níveis MT e BT das concessionárias brasileiras em cada *cluster* previamente formado.

4 APRESENTAÇÃO E ANÁLISE DOS RESULTADOS

Neste capítulo são apresentados e analisados inicialmente os resultados de perdas técnicas nos níveis de média e baixa tensão (MT e BT) para as 60 concessionárias de distribuição analisadas neste estudo.

Em seguida, expõem-se e discutem-se os resultados da clusterização executada com os três algoritmos disponíveis no *software SAS® Enterprise Guide 7.1*.

Finalizando o capítulo, é exibida a análise das variáveis e posterior aplicação do método de regressão *stepwise* para determinação dos parâmetros mais explicativos de perdas técnicas em MT e BT para cada *cluster* formado no processo anterior.

4.1 NÍVEIS DAS PERDAS TÉCNICAS NAS CONCESSIONÁRIAS DE DISTRIBUIÇÃO BRASILEIRAS

Conforme exposto no Capítulo 3, os valores de perdas técnicas em níveis de média e baixa tensão para as 60 concessionárias analisadas neste trabalho foram retirados das Notas Técnicas relativas ao Terceiro Ciclo de Revisões Tarifárias Periódicas, disponíveis no banco de dados da ANEEL e no sítio eletrônico da mesma.

A Tabela 4.1 mostra os percentuais de perdas técnicas resultantes da aplicação da metodologia para o cálculo das perdas nos sistemas de distribuição MT e BT adotada pela Agência conforme consta no Módulo 7 do PRODIST, Revisão 3 (05.04.2013).

Sobre as nomenclaturas utilizadas na Tabela 4.1, é válido frisar que a sigla “EnerTot_MTBT” faz menção à energia total injetada em SDMT e SDBT de cada concessionária no ano referente ao cálculo das perdas. Da mesma maneira, “Perdas_MTBT” é o montante de perdas nos níveis de média e baixa tensão calculadas para cada distribuidora no mesmo ano de referência. Finalmente, “PercPerdas_MTBT” é o percentual de perdas técnicas nesses níveis dado pela razão entre “Perdas_MTBT” e “EnerTot_MTBT”. A etapa de normalização das perdas técnicas pela energia total injetada é realizada com o objetivo de evitar que o processo de clusterização seja influenciado pelo tamanho das distribuidoras.

Tabela 4.1 - Percentuais de perdas técnicas em MT e BT das 60 distribuidoras analisadas.

Distribuidora	EnerTot_MTBT (MWh)	Perdas_MTBT (MWh)	PercPerdas_MTBT (%)
CPFL Jaguari	453.977,862	12.660,284	2,789%
EFLUL	93.031,370	3.079,242	3,310%
CPFL Piratininga	10.572.973,585	386.724,587	3,658%
EFLJC	13.117,342	505,196	3,851%
MUXENERGIA	61.703,049	2.391,356	3,876%
DMED	433.163,153	16.852,393	3,891%
ELETROPAULO	41.892.624,838	1.634.319,481	3,901%
AES-SUL	8.216.551,098	328.426,947	3,997%
CNEE	555.958,434	22.266,194	4,005%
ENF	351.510,589	14.340,582	4,080%
HIDROPAN	111.180,301	4.539,447	4,083%
EBO	620.588,265	26.207,629	4,223%
CPFL Paulista	25.607.313,622	1.091.920,159	4,264%
CPFL Santa Cruz	981.666,400	42.652,546	4,345%
CFLO	272.477,527	11.945,020	4,384%
AMPLA	11.671.496,390	530.763,105	4,548%
COCEL	282.078,113	13.064,270	4,631%
CELESC-Dis	17.624.286,164	828.796,112	4,703%
LIGHT	28.572.628,525	1.350.213,285	4,726%
RGE	9.316.898,425	448.428,259	4,813%
COPEL-Dis	23.684.794,869	1.143.467,526	4,828%
CAIUÁ-D	1.181.659,241	59.800,387	5,061%
EEB	759.831,242	38.614,818	5,082%
BANDEIRANTE	10.885.788,299	555.923,111	5,107%
CEB-Dis	6.076.646,547	323.867,090	5,330%
BOA VISTA	747.793,106	40.501,803	5,416%
CELPE	12.269.915,856	676.056,359	5,510%
CEEE-D	8.713.235,980	485.625,733	5,573%
ESE	2.599.034,628	145.689,555	5,606%
CEMAR	5.815.477,927	338.078,382	5,813%
EDEVP	796.058,695	46.781,990	5,877%
EMG	1.312.359,138	77.231,744	5,885%
DEMEI	137.957,930	8.221,903	5,960%
CEMIG-D	31.990.973,504	1.950.435,806	6,097%
COELCE	8.586.428,955	529.532,435	6,167%
COSERN	4.549.899,174	280.940,655	6,175%
COELBA	16.189.590,794	1.000.022,567	6,177%
AME	8.286.563,568	516.875,231	6,238%
ELETROCAR	177.147,028	11.369,656	6,418%
EPB	3.744.417,107	242.092,017	6,465%
ESCELSA	7.313.078,076	486.725,414	6,656%
ELFSM	457.983,477	30.568,360	6,675%
ELEKTRO	12.324.713,949	822.738,717	6,676%
EMS	4.970.801,731	332.646,430	6,692%
CELG-D	11.728.913,973	791.352,448	6,747%
CEAL	3.744.688,232	257.223,210	6,869%
CPFL Sul Paulista	441.161,994	30.964,589	7,019%
COOPERALIANÇA	192.280,868	14.021,633	7,292%
EMT	7.237.074,573	533.361,315	7,370%
CPFL Leste Paulista	341.381,761	25.533,097	7,479%
CPFL Mococa	226.552,646	19.182,761	8,467%
CEPISA	3.701.255,580	317.931,584	8,590%
ETO	1.670.003,528	144.555,116	8,656%
ELETROACRE	1.064.508,756	95.047,633	8,929%
CHESP	103.641,342	9.513,429	9,179%
CERON	3.547.243,136	327.757,541	9,240%
ENERGIA	242.663,498	23.313,911	9,608%
SULGIPE	296.602,057	29.708,060	10,016%
CELPA	8.355.148,838	878.950,991	10,520%
UHENPAL	73.654,656	8.723,620	11,844%

Nota-se na Tabela 4.1 que as empresas CPFL Jaguari, EFLUL, CPFL Piratininga, EFLJC, MUXENERGIA, DMED, ELETROPAULO e AES-SUL possuem os mais baixos percentuais de perdas técnicas, com índices menores que 4,000% em MT e BT. Por outro lado, as distribuidoras CHESP, CERON, IENERGIA, SULGIPE, CELPA e UHENPAL apresentam os mais elevados índices de perdas, ultrapassando 9,000%.

Também é possível observar na Tabela 4.1 que os níveis de perdas técnicas nas redes brasileiras de média e baixa tensão variam substancialmente para as 60 concessionárias analisadas, indo de 2,789% na CPFL Jaguari até 11,844% na UHENPAL.

Identifica-se ainda que o mercado e o tamanho das empresas de distribuição no Brasil são bastante díspares, dado que a menor concessionária em estudo apresenta uma injeção de energia total em SDMT e SDBT de 13.117,342 MWh (EFLJC), enquanto a maior possui uma injeção de 41.892.624,838 MWh (ELETROPAULO), valor quase 3200 vezes maior que o primeiro.

4.2 CLUSTERIZAÇÃO

Neste trabalho aplica-se a clusterização aglomerativa hierárquica para o agrupamento das empresas de distribuição no Brasil de acordo com seus percentuais de perdas técnicas em SDMT e SDBT.

A execução da clusterização nos resultados de perdas técnicas apresentados na Seção 4.1 permite agrupar concessionárias com níveis semelhantes de eficiência tocante a essas perdas na distribuição. Essa é a primeira etapa para a identificação dos parâmetros que mais explicam perdas em cada agrupamento de distribuidoras no país.

Nesta pesquisa, executa-se o processo de clusterização fazendo-se uso de três algoritmos de agrupamento:

- i. Média das distâncias entre objetos;
- ii. Distância dos centroides; e
- iii. Ward.

Em seguida, analisam-se seus respectivos resultados de forma a definir aquele que será empregado nas demais etapas deste estudo.

4.2.1 Definição do algoritmo a ser aplicado

Os resultados da clusterização das 60 distribuidoras com base nos percentuais de perdas técnicas em SDMT e SDBT são apresentados nas Tabelas 4.2, 4.3 e 4.4. Essas tabelas mostram os resultados dos agrupamentos realizados com os algoritmos de média das distâncias entre objetos, distância dos centroides e *Ward*, respectivamente.

Tabela 4.2 - R^2 e SR^2 da clusterização das 60 distribuidoras com base no percentual de perdas técnicas em níveis MT e BT – Algoritmo de média das distâncias.

Nº de Clusters	Clusters agrupados		SR^2	R^2
59	ELFSM	ELEKTRO	0,0000	1,0000
58	COSEEN	COELBA	0,0000	1,0000
57	ENF	HIDROPAN	0,0000	1,0000
56	AES-SUL	CNEE	0,0000	1,0000
55	EDEVP	EMG	0,0000	1,0000
54	COELCE	CL58	0,0000	1,0000
53	DMED	ELETROPAULO	0,0000	1,0000
52	RGE	COPEL-Dis	0,0000	1,0000
51	CL59	EMS	0,0000	1,0000
50	MUXENERGIA	CL53	0,0000	1,0000
49	CAIUÁ-D	EEB	0,0000	1,0000
48	CELESC-Dis	LIGHT	0,0000	1,0000
47	ESCELSA	CL51	0,0000	1,0000
46	CEEE-D	ESE	0,0000	1,0000
45	CL49	BANDEIRANTE	0,0000	1,0000
44	CPFLSantaCruz	CFLO	0,0000	1,0000
43	EFLJC	CL50	0,0000	1,0000
42	EBO	CPFLPaulista	0,0000	1,0000
41	ELETROCAR	EPB	0,0000	1,0000
40	CHESP	CERON	0,0000	1,0000
39	CL54	AME	0,0000	1,0000
38	CEPISA	ETO	0,0000	1,0000
37	CEMAR	CL55	0,0000	1,0000
36	CL47	CELG-D	0,0000	1,0000
35	COOPERALIANÇA	EMT	0,0000	1,0000
34	CL56	CL57	0,0000	1,0000
33	CELPE	CL46	0,0000	1,0000
32	COCEL	CL48	0,0000	1,0000
31	CEB-Dis	BOAVISTA	0,0000	1,0000
30	CEMIG-D	CL39	0,0000	1,0000

Nº de Clusters	Clusters agrupados		SR^2	R^2
29	CL37	DEMEI	0,0000	1,0000
28	CL42	CL44	0,0001	1,0000
27	CL32	CL52	0,0001	1,0000
26	CEAL	CPFLSulPaulista	0,0000	1,0000
25	CL35	CPFLLestePaulista	0,0001	0,9990
24	CPFLMococa	CL38	0,0001	0,9990
23	CL43	CL34	0,0002	0,9990
22	CL31	CL33	0,0002	0,9990
21	AMPLA	CL27	0,0001	0,9990
20	CL41	CL36	0,0004	0,9980
19	ELETROACRE	CL40	0,0002	0,9980
18	CL29	CL30	0,0008	0,9970
17	CPFLPiratiniga	CL23	0,0004	0,9970
16	CL20	CL26	0,0007	0,9960
15	CL21	CL45	0,0013	0,9950
14	CL17	CL28	0,0018	0,9930
13	IENERGIA	SULGIPE	0,0004	0,9930
12	CPFLJaguari	EFLUL	0,0006	0,9920
11	CL24	CL19	0,0020	0,9900
10	CL22	CL18	0,0044	0,9860
9	CL16	CL25	0,0048	0,9810
8	CL13	CELPA	0,0015	0,9800
7	CL14	CL15	0,0148	0,9650
6	CL10	CL9	0,0298	0,9350
5	CL11	CL8	0,0129	0,9220
4	CL12	CL7	0,0141	0,9080
3	CL4	CL6	0,2349	0,6730
2	CL5	UHENPAL	0,0270	0,6463
1	CL3	CL2	0,6463	0,0000

Tabela 4.3 - R^2 e SR^2 da clusterização das 60 distribuidoras com base no percentual de perdas técnicas em níveis MT e BT – Algoritmo de distância dos centroides.

Nº de Clusters	Clusters agrupados		SR^2	R^2
59	ELFSM	ELEKTRO	0,0000	1,0000
58	COSERN	COELBA	0,0000	1,0000
57	ENF	HIDROPAN	0,0000	1,0000
56	AES-SUL	CNEE	0,0000	1,0000
55	EDEVP	EMG	0,0000	1,0000
54	COELCE	CL58	0,0000	1,0000
53	DMED	ELETROPAULO	0,0000	1,0000
52	RGE	COPEL-Dis	0,0000	1,0000
51	CL59	EMS	0,0000	1,0000
50	MUXENERGIA	CL53	0,0000	1,0000
49	CAIUÁ-D	EEB	0,0000	1,0000
48	CELESC-Dis	LIGHT	0,0000	1,0000
47	ESCELSA	CL51	0,0000	1,0000
46	CEEE-D	ESE	0,0000	1,0000
45	CL49	BANDEIRANTE	0,0000	1,0000
44	EFLJC	CL50	0,0000	1,0000
43	CPFLSantaCruz	CFLO	0,0000	1,0000
42	EBO	CPFLPaulista	0,0000	1,0000
41	ELETROCAR	EPB	0,0000	1,0000
40	CHESP	CERON	0,0000	1,0000
39	CL54	AME	0,0000	1,0000
38	CEPISA	ETO	0,0000	1,0000
37	CEMAR	CL55	0,0000	1,0000
36	CL47	CELG-D	0,0000	1,0000
35	COOPERALIANÇA	EMT	0,0000	1,0000
34	CELPE	CL46	0,0000	1,0000
33	CL56	CL57	0,0000	1,0000
32	COCEL	CL48	0,0000	1,0000
31	CEB-Dis	BOAVISTA	0,0000	1,0000
30	CEMIG-D	CL39	0,0000	1,0000

Nº de Clusters	Clusters agrupados		SR^2	R^2
29	CL37	DEMEI	0,0000	1,0000
28	CL42	CL43	0,0001	1,0000
27	CL32	CL52	0,0001	1,0000
26	CL35	CPFLSulPaulista	0,0001	1,0000
25	CEAL	CPFLSulPaulista	0,0000	0,9990
24	CPFLMococa	CL38	0,0001	0,9990
23	CL44	CL33	0,0002	0,9990
22	CL31	CL34	0,0002	0,9990
21	AMPLA	CL27	0,0001	0,9990
20	CL41	CL36	0,0004	0,9980
19	ELETROACRE	CL40	0,0002	0,9980
18	CL29	CL30	0,0008	0,9970
17	CPFLPiratinga	CL23	0,0004	0,9970
16	CL20	CL25	0,0007	0,9960
15	CL21	CL45	0,0013	0,9950
14	CL17	CL28	0,0018	0,9930
13	IENERGIA	SULGIPE	0,0004	0,9930
12	CPFLJaguari	EFLUL	0,0006	0,9920
11	CL24	CL19	0,0020	0,9900
10	CL22	CL18	0,0044	0,9860
9	CL16	CL26	0,0048	0,9810
8	CL13	CELPA	0,0015	0,9800
7	CL14	CL15	0,0148	0,9650
6	CL10	CL9	0,0298	0,9350
5	CL11	CL8	0,0129	0,9220
4	CL12	CL7	0,0141	0,9080
3	CL4	CL6	0,2349	0,6730
2	CL5	UHENPAL	0,0270	0,6463
1	CL3	CL2	0,6463	0,0000

Tabela 4.4 - R^2 e SR^2 da clusterização das 60 distribuidoras com base no percentual de perdas técnicas em níveis MT e BT – Algoritmo de Ward.

Nº de Clusters	Clusters agrupados		SR^2	R^2
59	ELFSM	ELEKTRO	0,0000	1,0000
58	COSERN	COELBA	0,0000	1,0000
57	ENF	HIDROPAN	0,0000	1,0000
56	AES-SUL	CNEE	0,0000	1,0000
55	EDEVP	EMG	0,0000	1,0000
54	COELCE	CL58	0,0000	1,0000
53	DMED	ELETROPAULO	0,0000	1,0000
52	RGE	COPEL-Dis	0,0000	1,0000
51	CL59	EMS	0,0000	1,0000
50	CAIUÁ-D	EEB	0,0000	1,0000
49	CELESC-Dis	LIGHT	0,0000	1,0000
48	MUXENERGIA	CL53	0,0000	1,0000
47	ESCELSA	CL51	0,0000	1,0000
46	CEEE-D	ESE	0,0000	1,0000
45	CPFLSantaCruz	CFLO	0,0000	1,0000
44	CL50	BANDEIRANTE	0,0000	1,0000
43	EBO	CPFLPaulista	0,0000	1,0000
42	EFLJC	CL48	0,0000	1,0000
41	ELETROCAR	EPB	0,0000	1,0000
40	CHESP	CERON	0,0000	1,0000
39	CEPISA	ETO	0,0000	1,0000
38	COOPERALIANÇA	EMT	0,0000	1,0000
37	CEMAR	CL55	0,0000	1,0000
36	CL54	AME	0,0000	1,0000
35	AMPLA	COCEL	0,0000	1,0000
34	CEB-Dis	BOAVISTA	0,0000	1,0000
33	CL47	CELG-D	0,0000	1,0000
32	CELPE	CL46	0,0000	1,0000
31	CL56	CL57	0,0000	1,0000
30	CEMIG-D	CL36	0,0000	1,0000

Nº de Clusters	Clusters agrupados		SR^2	R^2
29	CL37	DEMEI	0,0000	1,0000
28	CEAL	CPFLSulPaulista	0,0000	1,0000
27	CL49	CL52	0,0001	1,0000
26	CL43	CL45	0,0001	1,0000
25	CL38	CPFLSulPaulista	0,0001	1,0000
24	CPFLMococa	CL39	0,0001	0,9990
23	CPFLPiratinga	CL42	0,0002	0,9990
22	CL35	CL27	0,0002	0,9990
21	CL34	CL32	0,0002	0,9990
20	ELETROACRE	CL40	0,0002	0,9990
19	IENERGIA	SULGIPE	0,0004	0,9980
18	CL41	CL33	0,0004	0,9980
17	CL23	CL31	0,0004	0,9970
16	CPFLJaguari	EFLUL	0,0006	0,9970
15	CL18	CL28	0,0007	0,9960
14	CL29	CL30	0,0008	0,9950
13	CL22	CL44	0,0013	0,9940
12	CL19	CELPA	0,0015	0,9930
11	CL17	CL26	0,0018	0,9910
10	CL24	CL20	0,0020	0,9890
9	CL21	CL14	0,0044	0,9840
8	CL15	CL25	0,0048	0,9800
7	CL16	CL11	0,0076	0,9720
6	CL12	UHENPAL	0,0107	0,9610
5	CL7	CL13	0,0213	0,9400
4	CL10	CL6	0,0291	0,9110
3	CL9	CL8	0,0298	0,8810
2	CL5	CL3	0,2349	0,6463
1	CL2	CL4	0,6463	0,0000

Da análise das Tabelas 4.2 e 4.3, percebe-se que o índice de homogeneidade R^2 é máximo (igual a 1,0000) até a formação de 25 *clusters* para a clusterização pela média das distâncias e para o algoritmo das distâncias dos centroides. Finalmente, para a clusterização com o algoritmo *Ward*, o parâmetro R^2 exibe o primeiro decréscimo, de 1,0000 para 0,9990, no estágio de redução de 25 para 24 *clusters*, indicando perda de homogeneidade.

Percebe-se também que, a partir do estágio de formação de 6 *clusters*, o R^2 apresenta redução mais intensa. Na formação de 3 *clusters*, os dois primeiros algoritmos resultam em valores de R^2 iguais a 0,6730, enquanto o algoritmo de *Ward* apresenta R^2 de 0,8810. Portanto, os resultados supracitados demonstram que a clusterização com base no percentual de perdas técnicas em níveis de média e baixa tensão possuem maior homogeneidade quando do emprego do método de *Ward*.

Conclui-se, portanto, que o algoritmo de *Ward* é o mais adequado para ser aplicado nos processos de clusterização realizados neste estudo.

4.2.2 Clusterização das distribuidoras com base nos percentuais das perdas técnicas em SDMT e SDBT

Dado o resultado da Seção 4.2.1, prossegue-se à análise da clusterização resultante do emprego do algoritmo de *Ward*. A Tabela 4.4 expõe os valores de homogeneidade R^2 e SR^2 do agrupamento das 60 concessionárias em análise com base nos seus percentuais de perdas técnicas em níveis de média e baixa tensão.

A Figura 4.1 apresenta o dendrograma do processo de clusterização com base no percentual de perdas técnicas em SDMT e SDBT. Nessa figura, o eixo horizontal contém a magnitude do SR^2 e o vertical apresenta as distribuidoras. As linhas verticais mais finas destacam os agrupamentos, e a vertical vermelha indica o valor de SR^2 no estágio de formação de 3 grupos (0,0298). As linhas horizontais apontam os *clusters* formados em cada etapa do processo.

Nota-se na Figura 4.1 que SR^2 permanece próximo a zero para agrupamentos de distribuidoras que possuem percentual de perdas técnicas próximos entre si. Os valores de SR^2 ficam mais elevados à medida que os grupos menores são aglomerados. Partindo dos agrupamentos CL7 e CL13, forma-se, primeiramente, um *cluster* maior (CL5), com SR^2 igual a 0,0076. Em seguida, CL6 e CL10 são agrupados formando o CL4, cujo SR^2 adquire o valor de 0,0291. Seguindo a ordem, CL8 e CL9 passam a formar o CL3, com SR^2 igual a 0,0298.

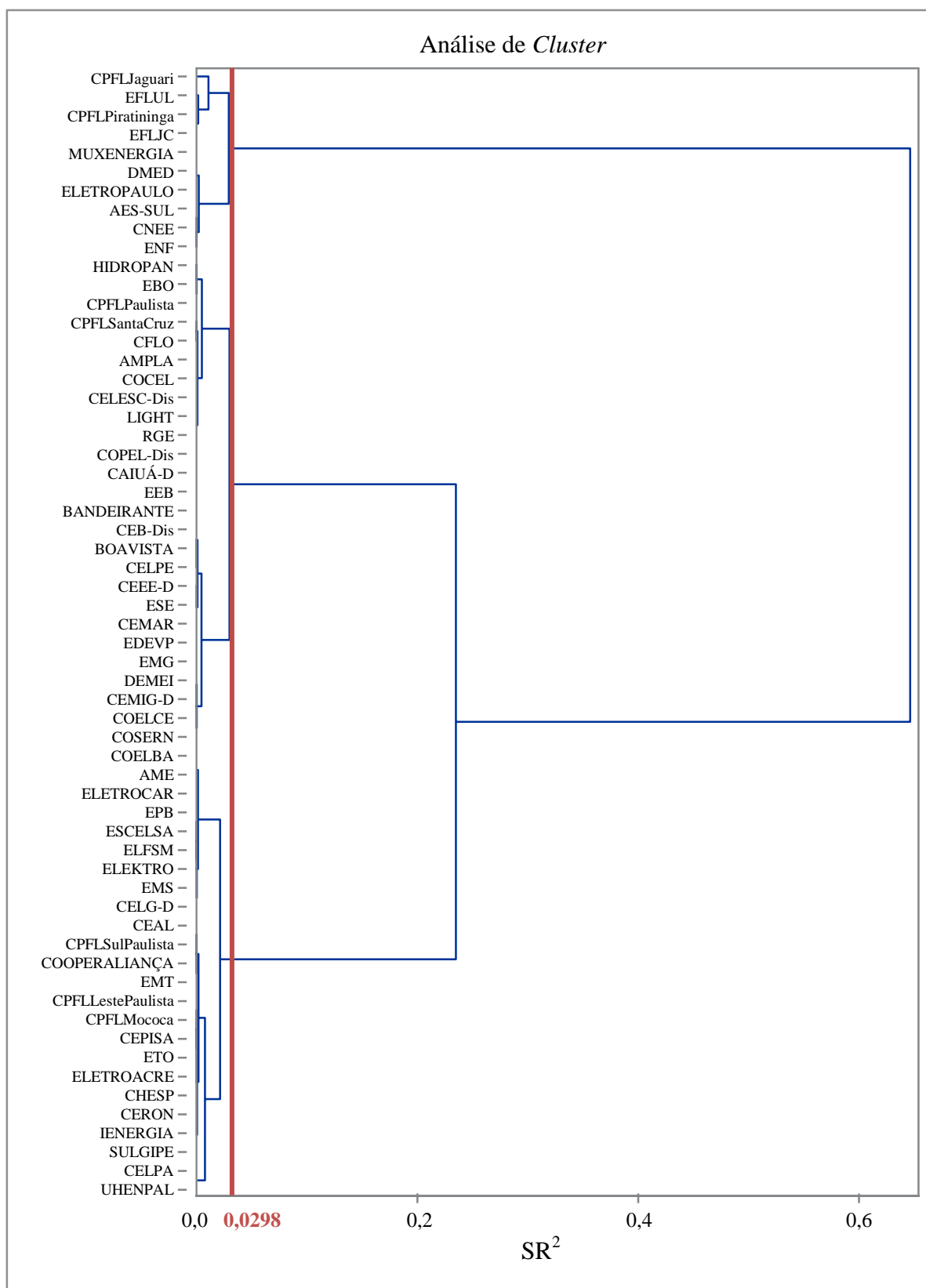


Figura 4.1 - Dendrograma da clusterização com base no percentual de perdas técnicas em níveis de média e baixa tensão.

É possível perceber ainda que o próximo passo aglomerativo culmina na fusão dos grupos CL3 e CL5, resultando no CL2 e em R^2 semi-parcial de 0,2349. Finalmente, CL2 e CL4 são aglomerados em um único *cluster*, cujo SR^2 apresenta o valor de 0,6463. Neste instante, como é possível extrair também da Tabela 4.4, R^2 passa a ser nulo, indicando homogeneidade mínima do agrupamento.

A Figura 4.2 apresenta a evolução do parâmetro SR^2 em relação à quantidade de *clusters* formados no processo de clusterização.

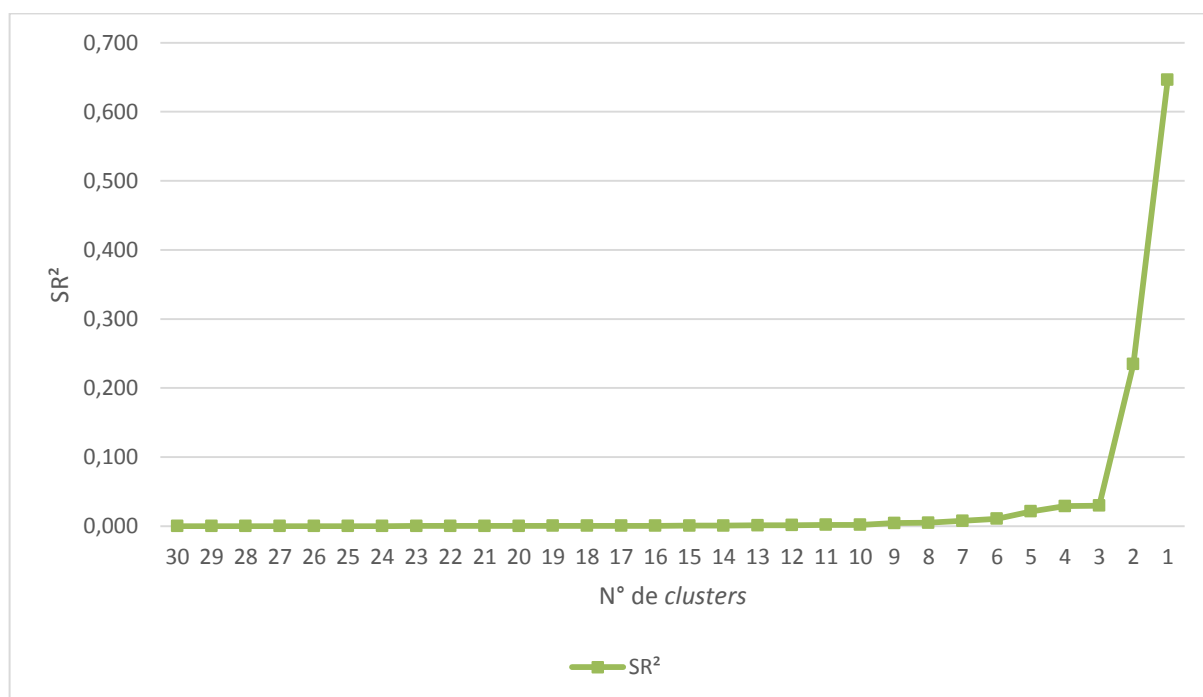


Figura 4.2 - Evolução do parâmetro SR^2 em relação à quantidade de *clusters* formados no processo de agrupamento.

Depreende-se da análise da Figura 4.2, que SR^2 se mantém muito próximo de zero até a constituição de 3 *clusters*, ocasião em que ele aumenta para próximo de 0,0298. Esse comportamento aponta para uma grande homogeneidade até essa etapa do processo de agrupamento. Nota-se que há, nos estágios seguintes, um crescimento acentuado desse parâmetro.

Portanto, conclui-se que, nas clusterizações executadas neste trabalho, a divisão das concessionárias em 3 grupos representa a menor quantidade de *clusters* com garantia de elevada homogeneidade entre seus componentes. Essa quantidade é utilizada como referência para as análises seguintes.

A Tabela 4.5 apresenta os três agrupamentos formados no processo de clusterização das distribuidoras com base no percentual de perdas técnicas em níveis de média e baixa tensão.

Tabela 4.5 - Grupos formados na clusterização das concessionárias com base no percentual de perdas técnicas em MT e BT.

<i>Cluster 1</i>		<i>Cluster 2</i>		<i>Cluster 3</i>	
Distribuidoras	Perc_PerdasMTBT	Distribuidoras	Perc_PerdasMTBT	Distribuidoras	Perc_PerdasMTBT
CPFL Jaguari	2,789%	CEB	5,330%	CPFL Mococa	8,467%
EFLUL	3,310%	BOA VISTA	5,416%	CEPISA	8,590%
CPFL Piratininga	3,658%	CELPE	5,510%	ETO	8,656%
EFLJC	3,851%	CEEE	5,573%	ELETROACRE	8,929%
MUXENERGIA	3,876%	ESE	5,606%	CHESP	9,179%
DMED	3,891%	CEMAR	5,813%	CERON	9,240%
ELETROPAULO	3,901%	EDEVP	5,877%	IENERGIA	9,608%
AES-SUL	3,997%	EMG	5,885%	SULGIPE	10,016%
CNEE	4,005%	DEMEI	5,960%	CELPA	10,520%
ENF	4,080%	CEMIG	6,097%	UHENPAL	11,844%
HIDROPAN	4,083%	COELCE	6,167%		
EBO	4,223%	COSERN	6,175%		
CPFL Paulista	4,264%	COELBA	6,177%		
CPFL Santa Cruz	4,345%	AME	6,238%		
CFLO	4,384%	ELETROCAR	6,418%		
AMPLA	4,548%	EPB	6,465%		
COCEL	4,631%	ESCELSA	6,656%		
CELESC	4,703%	ELFSM	6,675%		
LIGHT	4,726%	ELEKTRO	6,676%		
RGE	4,813%	EMS	6,692%		
COPEL	4,828%	CELG	6,747%		
CAIUÁ	5,061%	CEAL	6,869%		
EEB	5,082%	CPFL Sul Paulista	7,019%		
BANDEIRANTE	5,107%	COOPERALIANÇA	7,292%		
		EMT	7,370%		
		CPFL Leste Paulista	7,479%		

Ao comparar-se a Tabela 4.5 com os níveis de perdas técnicas das distribuidoras em SDMT e SDBT apresentados na Tabela 4.1, percebe-se que o *cluster 1* é formado pelas empresas com menores valores de PercPerdas_MTBT, cujos índices vão de 2,789% a 5,107%. Por outro lado, o grupo 3 apresenta as concessionárias com mais altos níveis de perdas, que vão de 8,467% a 11,844%. Por fim, o *cluster 2* contempla as empresas com percentuais de perdas intermediários: de 5,330% a 7,479%.

É conveniente frisar que cerca de 96% das empresas do *cluster 1*, i.e. com menor níveis de perdas técnicas, atuam nas regiões Sul e Sudeste brasileiras; os outros 4% são referentes à empresa EBO, cuja área de concessão encontra-se no Nordeste. Por outro lado,

60% das concessionárias pertencentes ao *cluster* 3 atendem as regiões Norte e Nordeste do país, com os piores índices de perdas técnicas em MT e BT.

4.3 INVESTIGAÇÃO DOS ATRIBUTOS MAIS EXPLICATIVOS DAS PERDAS TÉCNICAS EM NÍVEIS DE MÉDIA E BAIXA TENSÃO

4.3.1 Análise da distribuição das variáveis utilizadas para estudo das concessionárias brasileiras

Antes da aplicação da regressão *stepwise* para a determinação das variáveis mais influentes nos níveis de perdas técnicas das distribuidoras, faz-se necessário avaliar a distribuição dos dados relativos às empresas em cada *cluster*. Esta etapa visa verificar se o pressuposto de normalidade é atendido. As variáveis originais que não se encaixam na distribuição normal em nenhum desses testes são submetidas a uma das três transformações disponíveis para aproximar suas distribuições à normal, conforme exposto no Capítulo 3.

É importante ressaltar que a transformação dos dados pela aplicação das transformações supracitadas não causa prejuízo à análise dos resultados.

O grupo inicial de parâmetros analisados é exposto na Tabela 3.1, onde constam as 43 variáveis independentes estudadas e o percentual de perdas técnicas em MT e BT. Os resultados de distribuição obtidos desta análise para cada agrupamento mostraram-se idênticos para todos os três *clusters*. É apresentada, portanto, a Tabela 4.6, onde constam as 33 variáveis remanescentes no trabalho com as respectivas transformações após a avaliação de suas distribuições.

Tabela 4.6 - Variáveis resmanescentes na análise de seleção com as respectivas transformações.

Variáveis remanescentes no modelo		
$\ln(S_InstTot_TD)$	$\sqrt{\ln(Dens_Regul_UC)}$	$\ln(DensRed_B_Area)$
$\ln(S_InstMed_TD)$	$\ln(UtilizMed_Regul)$	$PercRede_B$
$\ln(Dens_SInstTD_Area)$	$\ln(P_MedTot_MT)$	$\ln(EnerTot_Forn)$
$\ln(Dens_SInstTD_Rede)$	$\ln(DensPot_MT_Rede)$	$\ln(DensEnerg_Forn_Area)$
$\ln(Dens_SInstTD_UC)$	$ResMed_Segm_MT$	$\ln(DensEnerg_Forn_Rede)$
$\ln(Dens_TD_Area)$	$\ln(CompTotMed_MT)$	$\ln(DensEnerg_Forn_UC)$
$\ln(Dens_TD_Rede)$	$PercRede_MT$	$\ln(AREAT)$
$\ln(Dens_TD_UC)$	$\ln(DensPot_B_Rede)$	$\ln(NUC)$
$\ln(N_Regul)$	$\ln(ResMed_Segm_B)$	$\ln(Dens_UC_Area)$
$\ln(S_InstTot_Regul)$	$\ln(ResMed_Ramal_B)$	$\ln(Dens_UC_Rede)$
$\sqrt{\ln(Dens_Regul_Rede)}$	$\ln(CompTotMed_B)$	$\ln(PercPerdas_MTBT)$

A análise da Tabela 4.6 permite concluir que a transformação logarítmica foi a mais utilizada para adequação da distribuição dos dados, uma vez que o logaritmo natural foi aplicado em 28 das 33 variáveis remanescentes. Por outro lado, somente dois parâmetros (Dens_Regul_Rede e Dens_Regul_UC) foram adequados à normalidade com a aplicação da raiz quadrada, e três permaneceram no modelo sem necessidade de transformação: ResMed_Segm_MT, PercRede_MT e PercRede_B. Frisa-se também que a aplicação do valor inverso não se mostrou profícua na base de dados com as 44 variáveis iniciais.

Percebe-se ainda que a variável dependente PercPerdas_MTB_T foi submetida à transformação logarítmica, de modo a apresentar distribuição normal. Sendo assim, a variável dependente a ser utilizada no modelo de regressão será $\ln(\text{PercPerdas_MTBT})$.

4.3.2 Análise de correlações

São avaliadas as correlações de *Pearson* entre as variáveis independentes e a variável dependente $\ln(\text{PercPerdas_MTBT})$ em cada um dos agrupamentos formados na etapa de clusterização. As Tabelas 4.7, 4.8 e 4.9 apresentam o valor de todas as correlações de *Pearson* entre as variáveis independentes e a dependente, bem como o valor da significância dessas correlações para os *clusters* 1, 2 e 3 respectivamente.

Tabela 4.7 - Correlação de Pearson entre as variáveis independentes e a dependente para o *cluster* 1.

Variável	Coefficiente de Correlação de Pearson	Prob > r para H0: Rho=0	Variável	Coefficiente de Correlação de Pearson	Prob > r para H0: Rho=0
$\ln(\text{S_InstMed_TD})$	-0,4075	0,0481	PercRede_MT	0,2100	0,3248
$\ln(\text{DensEnerg_Forn_Rede})$	-0,3843	0,0637	$\sqrt{\text{Dens_Regul_UC}}$	0,2493	0,2402
$\ln(\text{Dens_UC_Rede})$	-0,3694	0,0757	$\ln(\text{CompTotMed_MT})$	0,2855	0,1763
$\ln(\text{DensEnerg_Forn_Area})$	-0,3073	0,1441	$\ln(\text{DensPot_MT_Rede})$	0,2926	0,1653
$\ln(\text{Dens_UC_Area})$	-0,3030	0,1501	$\ln(\text{ResMed_Segm_B})$	0,3575	0,0864
$\ln(\text{DensRed_B_Area})$	-0,2429	0,2528	$\ln(\text{Dens_SInstTD_UC})$	0,3755	0,0706
$\ln(\text{Dens_SInstTD_Area})$	-0,2341	0,2710	$\ln(\text{EnerTot_Forn})$	0,3821	0,0654
PercRede_B	-0,2100	0,3248	$\ln(\text{NUC})$	0,3831	0,0646
$\ln(\text{Dens_SInstTD_Rede})$	-0,2024	0,3430	$\ln(\text{P_MedTot_MT})$	0,3834	0,0644
$\ln(\text{Dens_TD_Area})$	-0,1471	0,4929	$\ln(\text{S_InstTot_Regul})$	0,3845	0,0636
$\ln(\text{CompTotMed_B})$	-0,1112	0,6048	ResMed_Segm_MT	0,3943	0,0566
$\ln(\text{UtilizMed_Regul})$	-0,0181	0,9333	$\ln(\text{DensPot_B_Rede})$	0,4106	0,0463
$\ln(\text{DensEnerg_Forn_UC})$	-0,0124	0,9543	$\ln(\text{S_InstTot_TD})$	0,4437	0,0299
$\ln(\text{ResMed_Ramal_B})$	0,0361	0,8670	$\ln(\text{Dens_TD_UC})$	0,4850	0,0163
$\sqrt{\text{Dens_Regul_Rede}}$	0,1055	0,6236	$\ln(\text{N_Regul})$	0,4853	0,0162
$\ln(\text{Dens_TD_Rede})$	0,1114	0,6042	$\ln(\text{AreaT})$	0,5460	0,0058

Tabela 4.8 - Correlação de Pearson entre as variáveis independentes e a dependente para o *cluster* 2.

Variável	Coefficiente de Correlação de Pearson	Prob > r para H0: Rho=0	Variável	Coefficiente de Correlação de Pearson	Prob > r para H0: Rho=0
ln(DensEnerg_Forn_UC)	-0,5202	0,0064	ln(Dens_SInstTD_Rede)	-0,0767	0,7095
ln(DensEnerg_Forn_Rede)	-0,4257	0,0302	ln(ResMed_Segm_B)	-0,0612	0,7664
PercRede_B	-0,3570	0,0734	ln(S_InstMed_TD)	-0,0438	0,8320
ln(DensPot_MT_Rede)	-0,3346	0,0947	ln(N_Regul)	-0,0360	0,8614
ln(DensPot_B_Rede)	-0,2835	0,1605	ln(Dens_TD_Rede)	-0,0339	0,8695
ln(P_MedTot_MT)	-0,2493	0,2193	ln(AREAT)	-0,0225	0,9130
ln(EnerTot_Forn)	-0,2214	0,2772	sqrt(Dens_Regul_UC)	0,0152	0,9414
ln(DensRed_B_Area)	-0,2166	0,2879	ln(Dens_SInstTD_UC)	0,0241	0,9070
ln(DensEnerg_Forn_Area)	-0,2084	0,3070	sqrt(Dens_Regul_Rede)	0,0250	0,9036
ln(NUC)	-0,1302	0,5261	ln(Dens_TD_UC)	0,0534	0,7956
ln(S_InstTot_TD)	-0,1283	0,5324	ln(ResMed_Ramal_B)	0,0627	0,7608
ln(Dens_UC_Area)	-0,1002	0,6262	ln(CompTotMed_B)	0,0989	0,6309
ln(Dens_UC_Rede)	-0,0933	0,6504	ResMed_Segm_MT	0,1133	0,5817
ln(Dens_TD_Area)	-0,0928	0,6520	ln(CompTotMed_MT)	0,2220	0,2758
ln(Dens_SInstTD_Area)	-0,0915	0,6566	ln(UtilizMed_Regul)	0,2618	0,1964
ln(S_InstTot_Regul)	-0,0905	0,6600	PercRede_MT	0,3570	0,0734

Tabela 4.9 - Correlação de Pearson entre as variáveis independentes e a dependente para o *cluster* 3.

Variável	Coefficiente de Correlação de Pearson	Prob > r para H0: Rho=0	Variável	Coefficiente de Correlação de Pearson	Prob > r para H0: Rho=0
PercRede_MT	-0,5156	0,1271	ln(DensEnerg_Forn_Rede)	0,0362	0,9209
ln(DensPot_MT_Rede)	-0,3307	0,3506	ln(Dens_TD_Area)	0,0761	0,8345
ln(S_InstTot_TD)	-0,2950	0,4079	ln(S_InstTot_Regul)	0,0818	0,8223
ln(Dens_TD_UC)	-0,2919	0,4131	ln(DensEnerg_Forn_Area)	0,1057	0,7713
ln(EnerTot_Forn)	-0,2830	0,4282	ln(Dens_SInstTD_Area)	0,1314	0,7176
ln(P_MedTot_MT)	-0,2662	0,4572	ln(Dens_SInstTD_Rede)	0,1344	0,7112
ln(NUC)	-0,2605	0,4674	ln(Dens_UC_Area)	0,1461	0,6872
ln(AREAT)	-0,2186	0,5440	ln(Dens_UC_Rede)	0,1694	0,6399
ln(DensPot_B_Rede)	-0,1915	0,5961	ln(ResMed_Ramal_B)	0,1810	0,6168
ln(DensEnerg_Forn_UC)	-0,1697	0,6393	ln(ResMed_Segm_B)	0,2230	0,5357
ResMed_Segm_MT	-0,1522	0,6748	ln(DensRed_B_Area)	0,2272	0,5278
ln(Dens_TD_Rede)	-0,0777	0,8311	ln(CompTotMed_MT)	0,2417	0,5012
ln(Dens_SInstTD_UC)	-0,0405	0,9115	ln(S_InstMed_TD)	0,2628	0,4632
ln(CompTotMed_B)	-0,0341	0,9256	sqrt(Dens_Regul_UC)	0,4369	0,2068
ln(N_Regul)	-0,0125	0,9726	PercRede_B	0,5156	0,1271
ln(UtilizMed_Regul)	0,0309	0,9325	sqrt(Dens_Regul_Rede)	0,6713	0,0336

Da análise das Tabelas 4.7, 4.8 e 4.9, percebe-se que nenhum atributo independente apresenta módulo de correlação com ln(PercPerdas_MBTB) inferior a 0,0100. Portanto, não é excluída nenhuma variável com base em baixo relacionamento linear entre as variáveis independentes e a dependente.

Por outro lado, no primeiro *cluster*, as variáveis $\ln(\text{AREAT})$ e $\ln(\text{N_Regul})$ são as mais correlacionadas com o atributo dependente, com coeficientes iguais a 0,5460 e 0,4853 em módulo, respectivamente. No *cluster* 2, as maiores correlações são atribuídas a $\ln(\text{DensEnerg_Forn_UC})$ e $\ln(\text{DensEnerg_Forn_Rede})$, cujos valores foram 0,5202 e 0,4257 nessa ordem. Complementarmente, o terceiro agrupamento apresentou maior relacionamento linear entre $\ln(\text{PercPerdas_MTBT})$ e as variáveis $\sqrt{\text{Dens_Regul_Rede}}$ (coeficiente de *Pearson*: 0,6713), PercRede_B e PercRede_MT (coeficientes de *Pearson*: 0,5167 para os dois últimos atributos).

Conforme apresentado no Capítulo 3, também devem ser filtradas variáveis independentes muito correlacionadas entre si, dado que elas podem ser consideradas como representativas de uma mesma grandeza. Portanto, são avaliadas as correlações de *Pearson* entre todas as variáveis independentes.

Os atributos com correlação entre si maior que 0,9000 em módulo são considerados muito correlacionados. Sendo assim, para cada par com essa característica, a variável com maior correlação com a variável dependente $\ln(\text{PercPerdas_MTBT})$ é mantida e a outra, excluída. As Tabelas 4.10, 4.11 e 4.12 apresentam as variáveis mantidas e excluídas por esse critério para os três *clusters* em análise.

Tabela 4.10 - Variáveis independentes com correlações elevadas entre si e representativas de uma grandeza semelhante no *cluster* 1.

Variável Mantida	Variável Excluída	Correlação de <i>Pearson</i>
$\ln(\text{Dens_SInstTD_Area})$	$\ln(\text{Dens_TD_Area})$	0,9715
$\ln(\text{P_MedTot_MT})$	$\ln(\text{DensPot_MT_Rede})$	0,9668

Tabela 4.11 - Variáveis independentes com correlações elevadas entre si e representativas de uma grandeza semelhante no *cluster* 2.

Variável Mantida	Variável Excluída	Correlação de <i>Pearson</i>
$\ln\text{NUC}$	$\ln\text{S_InstTot_TD}$	0,9883
$\ln\text{S_InstTot_Regul}$	$\ln\text{N_Regul}$	0,9286

Tabela 4.12 - Variáveis independentes com correlações elevadas entre si e representativas de uma grandeza semelhante no *cluster* 3.

Variável Mantida	Variável Excluída	Correlação de Pearson
lnP_MedTot_MT	lnDensPot_B_Rede	0,9561
lnDens_SInstTD_Area	lnDens_TD_Area	0,9746
lnCompTotMed_MT	lnS_InstTot_Regul	0,9007
sqrtDens_Regul_Rede	sqrtDens_Regul_UC	0,9284
lnResMed_Segm_B	lnResMed_Ramal_B	0,9042

Das Tabelas 4.10, 4.11 e 4.12 é possível concluir que somente duas variáveis independentes são retiradas do modelo 1 nesta etapa, a saber: ln(Dens_TD_Area) e ln(DensPot_MT_Rede). Do agrupamento 2, são retirados também dois atributos: ln(S_InstTot_TD) e ln(N_Regul). O grupo em que há mais variáveis independentes com altos coeficientes de correlação de *Pearson* entre si é o terceiro, do qual são subtraídas cinco variáveis: ln(DensPot_B_Rede), ln(Dens_TD_Area), ln(S_InstTot_Regul), sqrt(Dens_Regul_UC) e ln(ResMed_Ramal_B).

Ao final do estudo das correlações, portanto, 30 variáveis independentes permaneceram no modelo dos *clusters* 1 e 2 para posterior análise de multicolinearidade. O modelo do *cluster* 3 conta com 27 atributos independentes para o estudo de correlação multivariada.

As Tabelas 4.13, 4.14 e 4.15 apresentam as variáveis remanescentes e utilizadas na próxima etapa de análise para cada agrupamento.

Tabela 4.13 - Variáveis dependentes remanescentes no modelo do *cluster* 1 após análise de correlação.

Variáveis Remanescentes no Cluster 1		
ln(S_InstTot_TD)	ln(UtilizMed_Regul)	PercRede_B
ln(S_InstMed_TD)	ln(P_MedTot_MT)	ln(EnerTot_Forn)
ln(Dens_SInstTD_Area)	ResMed_Segm_MT	ln(DensEnerg_Forn_Area)
ln(Dens_SInstTD_Rede)	ln(CompTotMed_MT)	ln(DensEnerg_Forn_Rede)
ln(Dens_SInstTD_UC)	PercRede_MT	ln(DensEnerg_Forn_UC)
ln(Dens_TD_Rede)	ln(DensPot_B_Rede)	ln(AREAT)
ln(Dens_TD_UC)	ln(ResMed_Segm_B)	ln(NUC)
ln(N_Regul)	ln(ResMed_Ramal_B)	ln(Dens_UC_Area)
ln(S_InstTot_Regul)	ln(CompTotMed_B)	ln(Dens_UC_Rede)
sqrt(Dens_Regul_Rede)	ln(DensRed_B_Area)	ln(PercPerdas_MTB_T)
sqrt(Dens_Regul_UC)		

Tabela 4.14 - Variáveis dependentes remanescentes no modelo do *cluster* 2 após análise de correlação.

Variáveis Remanescentes no <i>Cluster</i> 2		
ln(S_InstMed_TD)	ln(P_MedTot_MT)	PercRede_B
ln(Dens_SInstTD_Area)	ln(DensPot_MT_Rede)	ln(EnerTot_Forn)
ln(Dens_SInstTD_Rede)	ResMed_Segm_MT	ln(DensEnerg_Forn_Area)
ln(Dens_SInstTD_UC)	ln(CompTotMed_MT)	ln(DensEnerg_Forn_Rede)
ln(Dens_TD_Area)	PercRede_MT	ln(DensEnerg_Forn_UC)
ln(Dens_TD_Rede)	ln(DensPot_B_Rede)	ln(AREAT)
ln(Dens_TD_UC)	ln(ResMed_Segm_B)	ln(NUC)
ln(S_InstTot_Regul)	ln(ResMed_Ramal_B)	ln(Dens_UC_Area)
sqrt(Dens_Regul_Rede)	ln(CompTotMed_B)	ln(Dens_UC_Rede)
sqrt(Dens_Regul_UC)	ln(DensRed_B_Area)	ln(PercPerdas_MTBt)
ln(UtilizMed_Regul)		

Tabela 4.15 - Variáveis dependentes remanescentes no modelo do *cluster* 3 após análise de correlação.

Variáveis Remanescentes no <i>Cluster</i> 3		
ln(S_InstTot_TD)	ln(P_MedTot_MT)	ln(EnerTot_Forn)
ln(S_InstMed_TD)	ln(DensPot_MT_Rede)	ln(DensEnerg_Forn_Area)
ln(Dens_SInstTD_Area)	ResMed_Segm_MT	ln(DensEnerg_Forn_Rede)
ln(Dens_SInstTD_Rede)	ln(CompTotMed_MT)	ln(DensEnerg_Forn_UC)
ln(Dens_SInstTD_UC)	PercRede_MT	ln(AREAT)
ln(Dens_TD_Rede)	ln(ResMed_Segm_B)	ln(NUC)
ln(Dens_TD_UC)	ln(CompTotMed_B)	ln(Dens_UC_Area)
ln(N_Regul)	ln(DensRed_B_Area)	ln(Dens_UC_Rede)
sqrt(Dens_Regul_Rede)	PercRede_B	ln(PercPerdas_MTBt)
ln(UtilizMed_Regul)		

4.3.3 Avaliação da multicolinearidade das variáveis independentes

Conforme mencionado no Capítulo 2, a multicolinearidade é avaliada por meio da estatística denominada fator de inflação de variância (VIF). Valores elevados de VIF indicam que uma variável independente é bastante explicada pelas demais variáveis independentes presentes no modelo. Uma vez que essa relação tem como consequência a diminuição do poder preditivo dos atributos, busca-se minimizar a correlação multivariada na base de dados a ser utilizada para a regressão *stepwise*.

As Tabelas 4.16, 4.17 e 4.18 exibem as variáveis que foram retiradas do estudo por apresentarem valor de VIF superior a 5 nos *clusters* 1, 2 e 3 respectivamente.

Tabela 4.16 - Variáveis retiradas do modelo por apresentarem VIF maior que 5 no *cluster* 1.

Variáveis excluídas por VIF > 5	
ln(S_InstTot_TD)	ln(CompTotMed_MT)
ln(S_InstMed_TD)	ln(DensRed_B_Area)
ln(Dens_SInstTD_Area)	PercRede_B
ln(Dens_SInstTD_Rede)	ln(EnerTot_Forn)
ln(Dens_TD_Rede)	ln(DensEnerg_Forn_Area)
ln(Dens_TD_UC)	ln(DensEnerg_Forn_Rede)
ln(N_Regul)	ln(AREAT)
ln(S_InstTot_Regul)	ln(NUC)
sqrt(Dens_Regul_UC)	ln(Dens_UC_Area)
ln(P_MedTot_MT)	ln(Dens_UC_Rede)

Tabela 4.17 - Variáveis retiradas do modelo por apresentarem VIF maior que 5 no *cluster* 2.

Variáveis excluídas por VIF > 5	
ln(S_InstMed_TD)	PercRede_B
ln(Dens_SInstTD_Rede)	ln(EnerTot_Forn)
ln(Dens_TD_UC)	ln(DensEnerg_Forn_Rede)
sqrt(Dens_Regul_UC)	ln(AREAT)
ln(P_MedTot_MT)	ln(NUC)
PercRede_MT	ln(Dens_UC_Area)
ln(DensPot_B_Rede)	ln(Dens_UC_Rede)

Tabela 4.18 - Variáveis retiradas do modelo por apresentarem VIF maior que 5 no *cluster* 3.

Variáveis excluídas por VIF > 5	
ln(S_InstTot_TD)	ln(DensEnerg_Forn_Rede)
ln(P_MedTot_MT)	ln(DensEnerg_Forn_UC)
ln(Dens_SInstTD_Rede)	ln(AREAT)
ln(ResMed_Segm_B)	ln(NUC)
ln(EnerTot_Forn)	ln(Dens_UC_Area)
ln(N_Regul)	

A realização da avaliação da multicolinearidade das variáveis independentes resultou na retirada de 20, 14 e 11 atributos dos modelos dos agrupamentos 1, 2 e 3, nessa ordem. As variáveis que permaneceram no estudo após as investigações de distribuição, de correlação de *Pearson* e de multicolinearidade integram os modelos que serão levados à análise seguinte: a regressão *stepwise* para cada *cluster*.

4.3.4 Seleção de variáveis por meio de análise de regressão *stepwise*

O modelo *stepwise* deve ser estimado para cada um dos *clusters* formados pelo agrupamento das distribuidoras utilizando-se o algoritmo de *Ward*. A execução desta etapa permite

verificar quais parâmetros possuem maior influência no valor final de perdas técnicas em níveis de média e baixa tensão para cada agrupamento formado pelo processo de clusterização.

A regressão é realizada com apenas 11 variáveis remanescentes do primeiro *cluster* e 17 atributos do segundo e do terceiro *clusters*. As Tabelas 4.19, 4.20 e 4.21 exibem as variáveis utilizadas para a análise de regressão para os agrupamentos supracitados.

Tabela 4.19 - Variáveis utilizadas na regressão *stepwise* para o *cluster* 1.

Variáveis utilizadas na regressão <i>stepwise</i>	
ln(Dens_SInstTD_UC)	ln(ResMed_Segm_B)
sqrt(Dens_Regul_Rede)	ln(ResMed_Ramal_B)
ln(UtilizMed_Regul)	ln(CompTotMed_B)
ResMed_Segm_MT	ln(DensEnerg_Forn_UC)
PercRede_MT	ln(PercPerdas_MTBt)
ln(DensPot_B_Rede)	

Tabela 4.20 - Variáveis utilizadas na regressão *stepwise* para o *cluster* 2.

Variáveis utilizadas na regressão <i>stepwise</i> do <i>cluster</i> 2	
ln(Dens_SInstTD_Area)	ln(CompTotMed_MT)
ln(Dens_SInstTD_UC)	ln(ResMed_Segm_B)
ln(Dens_TD_Area)	ln(ResMed_Ramal_B)
ln(Dens_TD_Rede)	ln(CompTotMed_B)
ln(S_InstTot_Regul)	ln(DensRed_B_Area)
sqrt(Dens_Regul_Rede)	ln(DensEnerg_Forn_Area)
ln(UtilizMed_Regul)	ln(DensEnerg_Forn_UC)
ln(DensPot_MT_Rede)	ln(PercPerdas_MTBt)
ResMed_Segm_MT	

Tabela 4.21 - Variáveis utilizadas na regressão *stepwise* para o *cluster* 3.

Variáveis utilizadas na regressão <i>stepwise</i> do <i>cluster</i> 3	
ln(S_InstMed_TD)	ln(CompTotMed_MT)
ln(Dens_SInstTD_Area)	PercRede_MT
ln(Dens_SInstTD_UC)	ln(CompTotMed_B)
ln(Dens_TD_Rede)	ln(DensRed_B_Area)
ln(Dens_TD_UC)	PercRede_B
sqrt(Dens_Regul_Rede)	ln(DensEnerg_Forn_Area)
ln(UtilizMed_Regul)	ln(Dens_UC_Rede)
ln(DensPot_MT_Rede)	ln(PercPerdas_MTBt)
ResMed_Segm_MT	

O modelo *stepwise* é estimado considerando a variável $\ln(\text{PercPerdas_MTBT})$ como dependente. As etapas da seleção *stepwise* para cada *cluster* formado são apresentadas nas Tabelas 4.22, 4.23 e 4.24, onde também é mostrado o valor do coeficiente de determinação R^2 parcial (SR^2) para cada acréscimo de variáveis no modelo.

Tabela 4.22 - Etapas da seleção *stepwise* para o *cluster* 1.

Resumo da seleção <i>Stepwise</i>								
Passo	Variável Inserida	Variável Removida	#Variáveis no modelo	SR^2	R^2	C(p)	Valor F	Pr > F
1	$\ln(\text{DensPot_B_Rede})$		1	0,1686	0,1686	10,5881	4,46	0,0463
2	$\ln(\text{ResMed_Segm_B})$		2	0,1963	0,3648	5,3677	6,49	0,0188
3	ResMed_Segm_MT		3	0,0768	0,4416	4,5426	2,75	0,1128
4	$\ln(\text{Dens_SInstTD_UC})$		4	0,0736	0,5152	3,8349	2,88	0,1058
5	$\text{sqrt}(\text{Dens_Regul_Rede})$		5	0,0352	0,5504	4,5394	1,41	0,2505
6	$\ln(\text{DensEnerg_Forn_UC})$		6	0,0420	0,5925	4,9928	1,75	0,2029
7	$\ln(\text{ResMed_Ramal_B})$		7	0,0490	0,6415	5,1905	2,19	0,1587
8	$\ln(\text{CompTotMed_B})$		8	0,0035	0,6450	10,0966	0,14	0,7166
9	PercRede_MT		9	0,0026	0,6476	12	0,1	0,7612

Tabela 4.23 - Etapas da seleção *stepwise* para o *cluster* 2.

Resumo da seleção <i>Stepwise</i>								
Passo	Variável Inserida	Variável Removida	#Variáveis no modelo	SR^2	R^2	C(p)	Valor F	Pr > F
1	$\ln(\text{DensEnerg_Forn_UC})$		1	0,2706	0,2706	2,6609	8,9	0,0064
2	$\text{sqrt}(\text{Dens_Regul_Rede})$		2	0,0905	0,3611	1,2694	3,49	0,075
3	$\ln(\text{CompTotMed_MT})$		3	0,0401	0,4012	2,0582	1,69	0,2087
4	$\ln(\text{DensPot_MT_Rede})$		4	0,0415	0,4427	2,6562	1,82	0,1935
5	$\ln(\text{ResMed_Segm_B})$		5	0,0273	0,4700	3,7336	1,21	0,286
6	$\ln(\text{Dens_SInstTD_UC})$		6	0,0095	0,4795	5,4136	0,41	0,5327
7	ResMed_Segm_MT		7	0,0040	0,4835	7,2798	0,16	0,6933
8	$\ln(\text{DensEnerg_Forn_Area})$		8	0,0031	0,4866	11,0659	0,11	0,7418
9	$\ln(\text{ResMed_Ramal_B})$		9	0,0019	0,4885	13	0,07	0,8014
10		$\ln(\text{ResMed_Segm_B})$	8	0,0003	0,4882	11,0087	0,01	0,9269

Tabela 4.24 - Etapas da seleção *stepwise* para o *cluster* 3.

Resumo da seleção <i>Stepwise</i>								
Passo	Variável Inserida	Variável Removida	#Variáveis no modelo	SR^2	R^2	C(p)	Valor F	Pr > F
1	$\text{sqrt}(\text{Dens_Regul_Rede})$		1	0,4506	0,4506	.	6,56	0,0336
2	ResMed_Segm_MT		2	0,1453	0,5960	.	2,52	0,1566
3	$\ln(\text{Dens_TD_Rede})$		3	0,0669	0,6629	.	1,19	0,317
4	$\ln(\text{Dens_SInstTD_Area})$		4	0,1806	0,8435	.	5,77	0,0614
5	$\ln(\text{Dens_SInstTD_UC})$		5	0,0614	0,9050	.	2,59	0,1831
6	$\ln(\text{CompTotMed_MT})$		6	0,0502	0,9552	.	3,36	0,164
7	$\ln(\text{S_InstMed_TD})$		7	0,0270	0,9822	.	3,03	0,2238
8	PercRede_MT		8	0,0144	0,9966	.	4,17	0,29

Nas Tabelas 4.22, 4.23 e 4.24, o parâmetro SR^2 pode ser entendido como a contribuição adicional de cada variável independente para o poder de explicação da variável dependente. Sendo assim, conclui-se que os dois atributos que mais contribuem para a determinação das perdas técnicas em SDMT e SDBT nos *clusters* 1, 2 e 3 são, respectivamente:

- i) $\ln(\text{DensPot_B_Rede})$ e $\ln(\text{ResMed_Segm_B})$;
- ii) $\ln(\text{DensEnerg_Forn_UC})$ e $\text{sqrt}(\text{Dens_Regul_Rede})$; e
- iii) $\text{sqrt}(\text{Dens_Regul_Rede})$ e $\ln(\text{Dens_SInstTD_Area})$.

É interessante notar que o parâmetro $\text{sqrt}(\text{Dens_Regul_Rede})$ faz-se expressivo tanto no grupo 2 como no grupo 3.

Depreende-se da análise acima que, das 43 variáveis independentes iniciais, cinco são as mais relevantes para os resultados de perdas técnicas no cenário de distribuição de energia elétrica brasileiro. Faz-se necessário lembrar que o grupo 1 é formado pelas empresas com menores índices de perdas técnicas, enquanto o agrupamento 2 constitui-se das concessionárias com percentuais medianos dessas perdas. Por fim, as distribuidoras com maiores perdas foram aglomeradas no *cluster* 3.

A Tabela 4.25 apresenta as médias aritméticas do percentual de perdas técnicas em níveis de média e baixa tensão e de cada um dos parâmetros mais explicativos para os grupos 1, 2 e 3.

Tabela 4.25 - Médias aritméticas do percentual de perdas técnicas em níveis MT e BT e de cada um dos parâmetros mais explicativos para os *clusters* 1, 2 e 3.

	PercPerdas _MTBT	DensPot_B_Rede (MW/km)	ResMed_Segm_B (Ohms/km)	DensEnerg_Forn_UC (MWh/UC)	Dens_Regul_Rede (Regul/km)	Dens_SInstTD_Area (kVA/km ²)
Cluster 1	4,2564%	1349,9702	0,7446	2,2350	0,0112	311,0859
Cluster 2	6,3146%	787,9812	0,7899	1,8594	0,0095	119,0503
Cluster 3	9,5048%	312,1557	1,2164	1,6314	0,0064	26,6694

Nota-se na Tabela 4.25 que o percentual de perdas técnicas em SDMT e SDBT está positivamente relacionado com a resistência média do segmento em baixa tensão. Isto é, quanto menor o valor da resistência, menor é o nível de perdas. Esse resultado pode ser comprovado pela relação de perdas em linhas de distribuição por efeito *Joule* dada pela

expressão $P_{\text{perdas}} = R.I^2$, onde P_{perdas} é a potência dissipada na linha, R é a resistência do condutor, e I é a corrente que passa por esse condutor.

Por outro lado, também depreende-se da análise da Tabela 4.25 que PercPerdas_MTBT é inversamente proporcional a DensPot_B_Rede , DensEnerg_Forn_UC , Dens_Regul_Rede e Dens_SInstTD_Area . Esse resultado demonstra que o percentual das perdas técnicas em SDMT e SDBT são menores quanto maiores forem as densidades de potência por quilômetro de rede em baixa tensão, de energia fornecida anual por unidade consumidora, de reguladores por quilômetro de rede em média tensão e de potência nominal instalada em transformadores de distribuição por área atendida. Sendo assim, ao aumentarem-se as densidades supracitadas, melhora-se a eficiência da rede de distribuição quanto às perdas técnicas.

De fato, a densidade de potência instalada por quilômetro de rede em baixa tensão reflete o número de consumidores por extensão de rede nesse nível de tensão. Essa razão, bem como a intensidade do consumo dada pela densidade de energia fornecida por unidade consumidora por ano também foram apontadas como fatores influentes no resultado de perdas técnicas por Dortolina e Nadira (2005). Nesse estudo, os autores mostram a relação de dependência inversa que a densidade de consumidores por rede e a intensidade do consumo possuem com o percentual de perdas técnicas: quanto maiores os valores de consumidor/km e de kWh/consumidor/ano, menor é o nível dessas perdas percentuais no sistema.

Complementarmente, sabe-se que uma maior presença de reguladores na rede implica em um perfil de tensão mais estável por toda a distribuição. Níveis de tensão dentro de uma faixa estreita de operação e próximos a 1 pu implicam em correntes mais baixas para o atendimento das cargas conectadas ao sistema. Como exposto acima, as perdas por efeito *Joule* são menores para valores de correntes menos elevados.

Finalmente, a relação inversa entre o percentual de perdas técnicas e a densidade de potência nominal instalada em transformadores de distribuição por área atendida indica que uma área com maior densidade de consumo por km² implica em maior eficiência relativa às perdas técnicas na distribuição. É importante frisar que a existe também uma relação entre a capacidade de transformação e a tensão na linha, uma vez que, quanto menor a densidade de MW transformados por área, mais consumidores encontram-se distantes dos transformadores e, portanto, pior é o perfil de tensão na linha devido à queda na rede. Essa conclusão também

é sustentada pela retirada de Dens_SInstTD_Rede por conta da multicolinearidade, haja vista que Dens_SInstTD_Area traz a informação contida nesse outro atributo.

5 CONCLUSÕES

5.1 CONTRIBUIÇÕES DO TRABALHO

Este trabalho apresentou os resultados da análise de clusterização baseada nos percentuais de perdas técnicas em níveis de média e baixa tensão das distribuidoras brasileiras e posterior seleção das variáveis que mais explicam perdas para cada agrupamento formado. Tal análise foi realizada com os dados obtidos da base de dados da ANEEL referente aos indicadores coletivos de continuidade (DEC e FEC) e das Notas Técnicas relativas ao Terceiro Ciclo de Revisões Tarifárias Periódicas, realizado pela ANEEL entre os anos de 2011 e 2014.

A parte introdutória expôs a motivação para este trabalho, bem como o impacto que as perdas técnicas causam no custo da energia elétrica no Brasil. O capítulo 2 apresentou a metodologia utilizada pela ANEEL para o cálculo de perdas técnicas para o Terceiro Ciclo de Revisões Tarifárias Periódicas, como consta na Revisão 3 do Módulo 7 do PRODIST. Ainda nesse capítulo, discorreu-se sobre as técnicas utilizadas para avaliação das variáveis explicativas de perdas: i) clusterização e ii) análise de regressão.

No capítulo 3 foram descritos os materiais e métodos aplicados neste trabalho. Foram expostos nessa seção os dados empregados no estudo e os percentuais de perdas técnicas de cada uma das 60 distribuidoras brasileiras em análise, assim como os algoritmos de clusterização e as etapas da análise de variáveis utilizadas para a identificação dos parâmetros mais influentes para perdas técnicas em SDMT e SDBT por *cluster* formado. Dentre as etapas necessárias para a realização da regressão *stepwise* encontravam-se a avaliação da distribuição das variáveis independentes e dependente, a correlação de Pearson e a análise de multicolinearidade.

O capítulo 4 apresentou os resultados do estudo, por meio dos quais atesta-se que, para as distribuidoras com maior nível de eficiência relativa às perdas técnicas na distribuição, os fatores mais influentes nos seus percentuais de perdas técnicas são a densidade de potência por quilômetro de rede em baixa tensão e a resistência média do segmento no mesmo nível de tensão. Por outro lado, as concessionárias com níveis medianos de perdas possuem como maiores motivadores para os resultados de perdas as densidades de energia fornecida por unidade consumidora por ano e de reguladores por quilômetro de rede em média tensão. Este

último atributo também mostra-se expressivo nos resultados de perdas técnicas em SDMT e SDBT das empresas de distribuição com piores índices de eficiência no setor, ao lado da densidade de potência nominal instalada em transformadores de distribuição por área atendida.

Comparando-se os resultados das médias aritméticas das variáveis citadas acima para cada grupo de concessionárias, pôde-se concluir que, quanto menor o valor de resistência dos condutores em baixa tensão, menores são as perdas técnicas atribuídas à distribuição. Complementarmente, o resultado de perdas técnicas se mostra inversamente proporcional às densidades de potência por quilômetro de rede em baixa tensão, de energia fornecida por unidade consumidora por ano, de reguladores por quilômetro de rede em média tensão e de potência nominal instalada em transformadores de distribuição por área atendida. Isto é, quanto maior os valores desses atributos, menores são as perdas técnicas.

Relativamente à clusterização, mostrou-se que os maiores níveis de homogeneidade nos agrupamentos são atingidos com o algoritmo de *Ward* para a metodologia aglomerativa hierárquica. Sendo assim, esse algoritmo foi considerado o mais adequado para a utilização neste estudo. Com base nos resultados obtidos pelo emprego de *Ward*, definiu-se o agrupamento das concessionárias em 3 *clusters*.

Em face dos resultados apresentados, depreende-se que este estudo culminou em subsídios para a ANEEL e para agentes do setor elétrico brasileiro interessados em fomentar a eficiência no planejamento energético da distribuição das concessionárias do país.

5.2 SUGESTÃO DE TRABALHOS FUTUROS

Este trabalho utilizou os dados descritivos das redes de distribuição necessários para o cálculo de perdas técnicas para o Terceiro Ciclo de Revisões Tarifárias Periódicas, cuja metodologia consta na Revisão 3 do Módulo 7 do PRODIST. Atualmente, a ANEEL adota o método de fluxo de carga com o emprego de dados georreferenciados para o cálculo das perdas. As informações georreferenciadas têm o poder de representar os sistemas de distribuição brasileiros com maior acurácia em relação às redes reais. Propõe-se, portanto, que um novo estudo seja realizado com atributos retirados da Base de Dados Geográfica da Distribuidora (BDGD), adequando a análise das concessionárias de distribuição ao atual

cálculo de perdas realizado pela agência, conforme versa a Revisão 4 do mesmo módulo do PRODIST.

Adicionalmente, devido à insuficiência de informações, não foram contempladas neste trabalho a Companhia de Eletricidade do Amapá (CEA), a Companhia Energética de Roraima (CERR) e a empresa Força e Luz Coronel Vivida Ltda (FORCEL). Sugere-se que trabalhos futuros levantem as informações necessárias para que se torne possível a execução do estudo realizado para as demais concessionárias.

6 REFERÊNCIAS BIBLIOGRÁFICAS

ALMEIDA, Davi R. V. (2015). **Seleção de Variáveis para Análise Comparativa de Desempenho de Distribuidoras de Energia Elétrica**. Monografia de Pós-Graduação em Estatística Aplicada, Universidade Cruzeiro do Sul, Brasília, DF, 80p.

ANEEL (2013). **Procedimentos de Distribuição de Energia Elétrica no Sistema Elétrico Nacional – PRODIST: Módulo 7 – Cálculo de Perdas na Distribuição**. Revisão 3. Disponível em: http://www2.aneel.gov.br/arquivos/PDF/M%C3%B3dulo7_Revisao_4-2_.pdf. Acesso em: Setembro de 2017. Brasília, Brasil.

ANEEL (2016). **Notas Técnicas da Apuração das Perdas na Distribuição Referentes ao Terceiro Ciclo de Revisões Tarifárias Periódicas**. As informações relativas às perdas técnicas de energia das concessionárias atuantes no Brasil estão disponíveis na página eletrônica da ANEEL, por meio do seguinte endereço: <http://www2.aneel.gov.br/area.cfm?idArea=13>. Acesso em: Julho de 2016. Brasília, Brasil.

ANEEL (2017a). **BIG - Banco de Informação da Geração**. Disponível em: <http://www2.aneel.gov.br/aplicacoes/capacidadebrasil/capacidadebrasil.cfm>. Acesso em: Novembro de 2017. Brasília, Brasil.

ANEEL (2017b). **Consumidores, Consumo, Receita e Tarifa Média – Empresa, Classe de Consumo e Nível de Tensão**. Relatório do Sistema de Apoio à Decisão (SAD). Disponível em: <http://www2.aneel.gov.br/area.cfm?idArea=550>. Acesso em: Agosto de 2016. Brasília, Brasil.

ANEEL (2017c). **Aspectos Institucionais: Áreas de abrangência das concessionárias de distribuição de energia elétrica**. Disponível em: http://www2.aneel.gov.br/aplicacoes/atlas/aspectos_institucionais/2_4_1.htm. Acesso em: Novembro de 2017. Brasília, Brasil.

ANEEL (2017d). **Metodologia de Cálculo Tarifário da Distribuição: Perdas de Energia**. Disponível em: <http://www.aneel.gov.br/calculo-tarifario-e-metodologia>. Acesso em: Novembro de 2017. Brasília, Brasil.

DEVORE, Jay L. (2006). **Probabilidade e Estatística para Engenharia e Ciências**. 6ª edição. Pioneira Thomson Learning. São Paulo, Brasil.

DORTOLINA, Carlos A., NADIRA Ramón (2005). *The Loss That is Unkown is No Loss at All: A Top-Down/Bottom-Up Approach for Estimating Distribution Losses*. IEEE Transactions on Power Systems, vol. 20, no. 2, p. 1119-1125, maio de 2005. Estados Unidos.

EPE (2017). **Balanco Energético Nacional 2017**: Relatório Síntese - ano base 2016. Empresa de Pesquisa Energética – EPE.

EVERITT, Brian S. et al. (1980). *Cluster Analysis*. 5ª edição. John Wiley & Sons. Chichester, Inglaterra.

HAIR Jr, Joseph F. et al. (2007); trad. Adonai Schlup Sant’Anna e Anselmo Chaves Neto. **Análise Multivariada de Dados**. 5ª edição. Bookman. Porto Alegre, Brasil.

HARTIGAN, John A. (1975). *Clustering Algorithms*. John Wiley & Sons. Nova Iorque, Estados Unidos.

MINGOTI, Sueli A. (2005). **Análise de Dados Através de Métodos de Estatística Multivariada**: Uma Abordagem Aplicada. 1ª edição. Belo Horizonte: Editora UFMG.

MORRISON, Jeffrey S. (2006). **Seleção de Variáveis no Desenvolvimento de Modelos**. Revista Tecnologia de Crédito Serasa Experian. São Paulo, n. 57, p. 24-40, dezembro de 2006. Disponível em: http://www.serasaexperian.com.br/serasaexperian/publicacoes/revistas/2006/57/revista_0310.htm Acesso em: Outubro de 2017. Brasília, Brasil.

ONS (2016). **Plano da Operação Energética 2016/2020 - PEN 2016**. Operador Nacional do Sistema - ONS.

ONS (2017a). **Sobre o SIN**: Mapas. Disponível em: <http://ons.org.br/pt/paginas/sobre-o-sin/mapas>. Acesso em: Novembro de 2017. Brasília, Brasil.

ONS (2017b). **Sobre o SIN**: Sistemas Isolados. Disponível em: <http://ons.org.br/pt/paginas/sobre-o-sin/sistemas-isolados>. Acesso em: Novembro de 2017. Brasília, Brasil.

QUEIROZ, Leonardo M. O. (2010). **Estimação e Análise de Perdas Técnicas na Distribuição de Energia Elétrica**. Tese de Doutorado em Engenharia Elétrica, Departamento de Engenharia Elétrica, Universidade de Brasília, Brasília, DF, 155p.

SWINSCOW, T. D. V., CAMPBELL, M. J. (1997). *Statistics at Square One*. 9ª edição. Universidade de Southampton, Inglaterra.

TIMM, Neil H. (2002). *Applied Multivariate Analysis*. 1ª edição. Springer-Verlag. Nova Iorque, Estados Unidos.

VASCONCELOS, Marcos V. L. (2016). **Análise de custo-benefício da implantação das redes elétricas inteligentes nas concessionárias de distribuição de energia elétrica do Brasil**. Dissertação de Mestrado em Engenharia Elétrica, Publicação PPGEE.DM- XXX/16, Departamento de Engenharia Elétrica, Universidade de Brasília, Brasília, DF, 88p.

WORLD ENERGY COUNCIL (2017). *Energy Efficiency Indicators: Rate of electricity T&D losses*. Disponível em: <https://wec-indicators.enerdata.net/world-rate-of-electricity-T-D-losses.html>. Acesso em: Novembro de 2017. Brasília, Brasil.

XU, Rui; WUNSCH II, Donald C. (2009). *Clustering*. 1ª edição. Piscataway, NJ: IEEE Press. Estados Unidos.

APÊNDICES

APÊNDICE A Perdas na Distribuição

Este apêndice exibe os dados de balanço de energia e de perdas na distribuição para as 60 concessionárias analisadas neste trabalho. O levantamento dessas informações foi realizado com a consulta a todas as Notas Técnicas referentes ao Terceiro Ciclo de Revisões Tarifárias Periódicas realizado pela ANEEL.

Distribuidoras	Energia Total Injetada (MWh/ano)	Energia Total Fornecida (MWh/ano)	Perdas Técnicas (MWh/ano)	Perdas Não-Técnicas (MWh/ano)	Perdas Totais (MWh/ano)	Perdas Técnicas (%)	Perdas não-técnicas (%)	Perdas Totais (%)
CFLO	272.875,00	264.395,50	12.599,50	-4.120,00	8.479,50	4,617%	-1,510%	3,107%
EFLUL	93.031,37	89.355,82	3.162,55	513,00	3.675,55	3,399%	0,551%	3,951%
CPFL Jaguari	504.903,87	482.977,74	15.842,08	6.084,04	21.926,12	3,138%	1,205%	4,343%
EFLJC	13.117,34	12.441,50	531,88	143,96	675,84	4,055%	1,097%	5,152%
MUXENERGIA	61.703,05	58.377,80	2.478,82	846,44	3.325,25	4,017%	1,372%	5,389%
DMED	438.292,36	413.316,48	22.045,54	2.930,34	24.975,88	5,030%	0,669%	5,698%
EEB	1.054.575,71	992.358,49	52.380,38	9.836,85	62.217,22	4,967%	0,933%	5,900%
COCEL	283.418,83	266.628,78	14.815,20	1.974,85	16.790,04	5,227%	0,697%	5,924%
HIDROPAN	111.549,00	104.363,00	5.662,52	1.523,48	7.186,00	5,076%	1,366%	6,442%
ENF	365.080,94	339.973,88	19.748,75	5.358,31	25.107,06	5,409%	1,468%	6,877%
CPFL- Piratininga	17.123.939,74	15.855.795,87	732.797,73	535.346,14	1.268.143,86	4,279%	3,126%	7,406%
CELESC-Dis	21.688.987,00	20.029.701,25	1.375.070,13	284.215,61	1.659.285,74	6,340%	1,310%	7,650%
ELEKTRO	16.163.081,39	14.900.426,82	1.026.166,18	236.488,38	1.262.654,56	6,349%	1,463%	7,812%
EBO	676.520,34	619.834,70	30.523,54	26.162,10	56.685,64	4,512%	3,867%	8,379%
COPEL-Dis	28.487.907,45	26.095.500,22	1.808.390,98	584.016,31	2.392.407,30	6,348%	2,050%	8,398%
EDEVP	810.713,55	740.456,90	59.313,35	10.943,30	70.256,65	7,316%	1,350%	8,666%
CNEE	567.121,97	516.897,44	39.602,77	10.621,77	50.224,53	6,983%	1,873%	8,856%
CAIUA-D	1.191.340,57	1.085.466,99	77.281,68	28.591,91	105.873,58	6,487%	2,400%	8,887%
ELETROCAR	177.747,51	161.685,15	12.786,25	3.276,11	16.062,36	7,193%	1,843%	9,037%
CPFL Santa Cruz	1.060.287,85	960.653,17	83.180,01	16.454,67	99.634,68	7,845%	1,552%	9,397%
CPFL-Paulista	32.551.358,66	29.450.063,95	2.017.740,73	1.083.554,00	3.101.294,73	6,199%	3,329%	9,527%
COOPERALIANÇA	193.450,00	174.710,00	15.594,54	3.145,46	18.740,00	8,061%	1,626%	9,687%
CPFL Sul Paulista	507.353,90	457.833,98	33.968,46	15.551,46	49.519,92	6,695%	3,065%	9,760%
EMG	1.692.972,41	1.527.286,92	140.855,84	24.829,65	165.685,49	8,320%	1,467%	9,787%
RGE	10.431.907,98	9.386.881,27	752.535,21	292.491,51	1.045.026,72	7,214%	2,804%	10,018%
CPFL Mococa	229.113,01	205.496,28	21.746,69	1.870,05	23.616,74	9,492%	0,816%	10,308%
AES-SUL	9.702.860,16	8.675.485,61	604.819,04	422.555,51	1.027.374,55	6,233%	4,355%	10,588%
ELETROPAULO	48.856.816,80	43.496.601,76	2.544.851,97	2.815.363,11	5.360.215,08	5,209%	5,762%	10,971%
BANDEIRANTE	16.168.768,49	14.360.662,91	791.881,66	1.016.223,93	1.808.105,59	4,898%	6,285%	11,183%
ESE	3.544.981,10	3.143.215,66	263.982,06	137.783,38	401.765,44	7,447%	3,887%	11,333%
CPFL Leste Paulista	344.233,91	303.544,10	27.892,71	12.797,10	40.689,81	8,103%	3,718%	11,820%
CEB-Dis	6.707.339,85	5.914.248,38	478.992,18	314.099,30	793.091,48	7,141%	4,683%	11,824%
DEMEI	137.957,93	121.538,07	8.585,72	7.834,14	16.419,86	6,223%	5,679%	11,902%
ELFSM	462.000,92	406.852,72	39.513,37	15.634,83	55.148,20	8,553%	3,384%	11,937%
CEMIG-D	50.426.705,55	44.287.688,80	3.950.738,69	2.188.278,14	6.139.016,83	7,835%	4,340%	12,174%
COELCE	9.908.841,00	8.696.493,26	874.706,52	337.641,22	1.212.347,74	8,828%	3,407%	12,235%
COSERN	5.431.541,56	4.765.075,67	475.173,70	191.292,20	666.465,90	8,748%	3,522%	12,270%
CHESP	104.752,68	91.770,37	11.096,94	1.885,36	12.982,30	10,593%	1,800%	12,393%
BOA VISTA	823.145,84	719.894,34	57.967,61	45.283,89	103.251,51	7,042%	5,501%	12,544%
CELG-D	13.230.270,20	11.474.699,87	1.238.761,10	516.809,23	1.755.570,33	9,363%	3,906%	13,269%
IENERGIA	242.663,50	209.838,37	23.639,99	9.185,15	32.825,13	9,742%	3,785%	13,527%
ESCELSA	11.735.123,75	10.107.239,50	775.620,34	852.263,92	1.627.884,26	6,609%	7,263%	13,872%
UHENPAL	73.654,66	63.111,02	8.913,12	1.630,52	10.543,64	12,101%	2,214%	14,315%
SULGIPE	360.207,17	307.026,54	34.609,23	18.571,40	53.180,63	9,608%	5,156%	14,764%
ETO	1.782.437,58	1.518.827,36	236.336,33	27.273,89	263.610,22	13,259%	1,530%	14,789%
EPB	4.622.793,93	3.903.201,69	429.935,25	289.656,99	719.592,24	9,300%	6,266%	15,566%
COELBA	18.970.793,16	15.949.251,18	1.760.968,67	1.260.573,32	3.021.541,99	9,283%	6,645%	15,927%
EMT	7.968.175,98	6.527.191,86	789.450,24	651.533,89	1.440.984,13	9,908%	8,177%	18,084%
CEEE-D	9.913.206,90	7.991.398,84	641.279,29	1.280.528,78	1.921.808,07	6,469%	12,917%	19,386%
CELPE	14.339.194,75	11.508.919,67	1.161.904,95	1.668.370,14	2.830.275,08	8,103%	11,635%	19,738%
AMPLA	13.703.006,08	10.950.484,83	1.121.050,58	1.631.470,68	2.752.521,26	8,181%	11,906%	20,087%
EMS	5.274.308,85	4.126.872,64	648.200,02	499.236,18	1.147.436,21	12,290%	9,465%	21,755%
CEMAR	6.108.634,46	4.733.846,42	602.236,70	772.551,34	1.374.788,04	9,859%	12,647%	22,506%
ELETOACRE	1.067.890,60	822.431,67	105.188,45	140.270,47	245.458,92	9,850%	13,135%	22,985%
LIGHT	36.635.643,51	27.704.763,46	1.976.815,80	6.954.064,26	8.930.880,06	5,396%	18,982%	24,378%
CERON	3.637.302,01	2.629.173,34	405.469,97	602.658,70	1.008.128,67	11,148%	16,569%	27,716%
CEAL	4.325.513,00	3.102.547,08	447.422,56	775.543,36	1.222.965,92	10,344%	17,930%	28,273%
CEPISA	3.943.648,00	2.800.906,00	480.055,12	662.686,89	1.142.742,00	12,173%	16,804%	28,977%
CELPA	9.129.155,20	6.212.227,05	1.110.467,34	1.806.460,81	2.916.928,15	12,164%	19,788%	31,952%
AME	9.323.191,92	5.605.433,78	715.651,95	3.002.106,21	3.717.758,16	7,676%	32,200%	39,876%

A imagem a seguir expõe graficamente os dados referentes às perdas técnica e não técnica de todas as concessionárias em estudo.

